Graduate Texts in Mathematics

Marc Hindry
Joseph H. Silverman

Diophantine Geometry

An Introduction

Springer

Graduate Texts in Mathematics 201

Springer Science+Business Media, LLC

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces. 2nd ed.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy

(continued after index)

Marc Hindry
Joseph H. Silverman

Diophantine Geometry

An Introduction

Springer

Marc Hindry
Département de Mathématiques
Université Denis Diderot Paris 7
75251 Paris
France
hindry@math.jussieu.fr

Joseph H. Silverman
Department of Mathematics
Brown University
Providence, RI 02912
USA
jhs@math.brown.edu

With 8 illustrations.

Mathematics Subject Classification (1991): 11Gxx, 14Gxx

Library of Congress Cataloging-in-Publication Data
Hindry, Marc.
Diophantine geometry : an introduction / Marc Hindry, Joseph H. Silverman.
p. cm. — (Graduate texts in mathematics ; 201)
Includes bibliographical references and index.
ISBN 978-0-387-98981-5 ISBN 978-1-4612-1210-2 (eBook)
DOI 10.1007/978-1-4612-1210-2
1. Arithmetical algebraic geometry. I. Silverman, Joseph H., 1955– II. Title. III. Series.
QA242.5.H56 2000
512'.7—dc21 99-057467

Printed on acid-free paper.

Production managed by MaryAnn Brickner; manufacturing supervised by Jeffrey Taub.
Typeset by the authors using *Textures*.

9 8 7 6 5 4 3 2 1

ISBN 978-0-387-98981-5

To

Ramani, Natalia, Ariadne

and

Debby, Daniel, Jonathan

Preface

Number theory begins with the integers

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

and their natural operations:

addition and multiplication.

Functions $f : \mathbb{Z}^n \to \mathbb{Z}$ formed using only addition and multiplication are polynomial functions; so a system of polynomial equations

$$f_j(X_1, X_2, \ldots, X_m) = 0, \qquad 1 \le j \le m, \qquad (*)$$

to be solved in integers $X_1, \ldots, X_m \in \mathbb{Z}$ reflects the innermost structure of the fundamental operations, addition and multiplication, performed on the most fundamental of mathematical objects, the integers. Systems of polynomial equations $(*)$ are called *Diophantine equations* after Diophantus ($\Delta\iota\acute{o}\phi\alpha\nu\tau o\varsigma$) of Alexandria, ca. A.D. 100, whose *Arithmetica* contains numerous solved problems of this type.

From a geometric perspective, the complex solutions to a system of equations such as $(*)$ form an algebraic variety. Algebraic geometry, the study of algebraic varieties, also has a long and honorable history, although not quite as venerable as that of number theory. During the course of the 20th century it became clear that the deep and powerful concepts and methods of algebraic geometry are ideal for the study of Diophantine equations. This led Serge Lang in 1961 to coin the phrase "Diophantine Geometry" for the title of a book in which he sought to exploit the most powerful techniques of algebraic geometry to study Diophantine equations in their most general setting.

Later in the century, the field of algebraic geometry itself was reformulated by the Grothendieck school in such a way that one might plausibly argue that number theory, or at least the theory of Diophantine equations, is simply the special case of algebraic geometry over the spectrum of a

Dedekind domain! This view of number theory, sometimes dubbed "arithmetic geometry," has enjoyed considerable success, but to fully understand and exploit its power requires a substantial background in Grothendieck-style algebraic geometry, with the material in a book such as Hartshorne [1] providing a bare beginning.

In this volume we study Diophantine equations using tools from algebraic number theory and "classical" algebraic geometry. Our goal is to prove four of the fundamental finiteness theorems in Diophantine geometry:

★ Mordell–Weil Theorem
The group of rational points on an abelian variety is finitely generated.

★ Roth's Theorem
An algebraic number has finitely many approximations of order $2+\epsilon$.

★ Siegel's Theorem
An affine curve of genus $g \geq 1$ has finitely many integral points.

★ Faltings' Theorem
A curve of genus $g \geq 2$ has finitely many rational points.

We have chosen to avoid the use of scheme-theoretic language and concepts in our main development and in our proofs, so as to make the results more easily accessible, but we do include a substantial amount of supplementary material of a more advanced nature, usually without proof. We have also included a lengthy introduction to algebraic geometry in Part A, since our experience is that the real conundrum for students attempting to study Diophantine geometry is how to acquire a sufficient grasp of algebraic geometry without first spending years on purely geometric study.

The last decade of the 20th century saw explosive progress in the study of Diophantine geometry, but each major advance serves to highlight just how little we know and how much is left to discover and to prove. It is our hope that this book will help you, the reader, to appreciate some of the deep and elegant Diophantine results currently known and will inspire you to add to our knowledge of this beautiful subject.

Acknowledgments

The writing of this book has occupied the better part of a decade, and each author has taught several courses based on the material contained herein. It is thus impossible at this late date for us to accurately catalog the numerous colleagues, students, and friends who have looked at various drafts and offered suggestions, corrections, and helpful criticisms, but to all of you we offer our gratitude and acknowledge our great debt. We would like in particular to thank Hervé Billard, Florian Breuer, Antoine Chambert-Loir, Laurent Denis, Teresa de Diego, Martine Girard, Paul Lockhart, Sandra Marcello, Andrea Surroca, and Siman Wong for their assistance. We also want to express our appreciation to the many people from whom we learned

this subject, including (but certainly not limited to) Serge Lang, Jean-Pierre Serre, Lucien Szpiro, and John Tate.

In writing this volume we have consulted a great many sources. We have tried to provide citations for major theorems, but many results that are now considered "standard" have been presented as such. In any case, we claim no originality for any of the unlabeled material in this book and apologize in advance to anyone who feels slighted. Sources that we found especially useful include Bombieri [1], Lang [6], Schmidt [1], and Serre [3]. We would also like to thank Professor Bombieri for his permission to include a lengthy quotation from his article [1] in Section E.3.

Finally, and most importantly, we want to thank our wives and our offspring for their love and support and for providing all of those wonderful distractions that help to remind us that there is more to life than mathematics.

Marc Hindry
Joseph H. Silverman
January 1, 2000

Contents

PART A—DETAILED CONTENTS

Introduction

Diophantine equations are systems of polynomial equations to be solved in integers or rational numbers, and Diophantine geometry is the study of Diophantine equations using ideas and techniques from algebraic geometry. This is a very natural approach, since the basic objects studied in algebraic geometry, namely algebraic varieties, are themselves defined by systems of polynomial equations. The difference is that a (classical) algebraic geometer studies solutions in complex numbers or in some other algebraically closed field, while a number theorist studies solutions in a ring or field of arithmetic interest.

The most obvious way to classify polynomial equations is by their degrees, but whether one studies algebraic geometry or Diophantine equations, it soon becomes clear that such a classification is sadly lacking. For example, the equations

$$V_1 : y^2 = x^5 + x^4 \qquad \text{and} \qquad V_2 : y^2 = x^5 + x$$

appear similar, but V_1 has infinitely many solutions in rational numbers x and y, while V_2 has only finitely many such solutions, and indeed has only finitely many solutions with x and y chosen from any number field. One is inexorably led to search for more intrinsic invariants. For curves such as V_1 and V_2, the desired quantity is the genus; more generally, useful invariants may be defined using, for example, sheaves of differentials. In any case, one attempts to classify varieties geometrically according to various discrete and/or continuous parameters, and to describe the parameter spaces, which themselves often turn out to be varieties.

The geometric classification of curves (i.e., irreducible varieties of dimension 1) is extremely easy to describe. First, every curve is birational to a unique nonsingular projective curve. Second, every such curve has associated to it a nonnegative integer called its genus. Third, the isomorphism classes of nonsingular projective curves of a given genus g form (in a certain complicated, but well-defined, sense) a family of dimension $\max\{3g-3, g\}$. The proof of these assertions is not difficult.

The arithmetic classification of curves is almost as easy to describe, but many of the proofs lie very deep and will be our principal concern

for much of this volume. To simplify matters, we will assume that the curve C is projective and nonsingular of genus g, and that it has at least one rational point. Then the fundamental Diophantine finiteness theorems for curves can be summarized in the following short table.

Genus	Points "at Infinity"	Integer Solutions
$g = 0$	≤ 1	infinite set
$g = 0$	$= 2$	finitely generated group
$g = 0$	≥ 3	finite set
$g = 1$	$= 0$	finitely generated group
$g = 1$	≥ 1	finite set
$g \geq 2$	≥ 0	finite set

The Arithmetic Classification of Curves

(By convention, if a curve or equation has zero points at infinity, then its integer solutions coincide with its solutions in rational numbers.) If we introduce the Euler characteristic

$$\chi(C) = 2 - 2g - (\#\text{ of points "at infinity"}),$$

then the above results take the strikingly simple form

Euler Characteristic	Integer Solutions
$\chi(C) > 0$	infinite set
$\chi(C) = 0$	finitely generated group
$\chi(C) < 0$	finite set

This innocuous little table includes major theorems associated with the names Dirichlet, Mordell, Siegel, Weil, and Faltings.

A fundamental lesson to be learned from the above table is that at least for curves, and at least in a qualitative sense,

Geometry Determines Arithmetic

This, then, is the principal motivation and ultimate goal of Diophantine geometry—to describe the solutions of systems of Diophantine equations in terms of the geometric properties and invariants of the associated algebraic varieties. For curves, this task has been largely completed at the qualitative level, although there are many questions of a more refined nature that remain unanswered. For surfaces and varieties of higher dimension, the task is barely begun, and indeed in many cases the "right" conjectures have only recently been or are yet to be formulated.

Our study of Diophantine geometry begins in Part A with geometry. We give an overview of the algebro-geometric material that will be used in the rest of the book. This part discusses algebraic varieties, divisors, linear systems, algebraic curves (and surfaces), abelian varieties, Jacobian varieties, and schemes. Virtually all of this material, other than the section on schemes, is used in our subsequent work, but this does not mean that we recommend reading Part A in full before proceeding to the later parts of the book. Instead, we suggest briefly looking over Part A to see what it contains, and then jumping directly into the arithmetic material of Part B and beyond. Then return to Part A to fill in the algebraic geometry as it is needed in the later sections of the book.

The first arithmetic portion of the book, Part B, deals with the theory of height functions. These are functions that on the one hand measure the arithmetic complexity of a point on a variety, and on the other hand satisfy nice geometric transformation laws. Briefly, the theory of height functions is a tool that transforms geometric facts into number theoretic facts. More precisely, it transforms a divisor relation into an arithmetic complexity relation. These arithmetic relations are, in general, given only up to undetermined bounded quantities, but on abelian varieties it is possible to pick out particular height functions, called canonical heights, for which the arithmetic relations become exact. The material on heights and canonical heights required in Parts C–E is covered in Sections B.1–B.6. Section B.7 contains some useful lemmas used in subsequent sections, and Sections B.8–B.10 describe further important topics, often without proof, that are used only in Part F.

We then come to the Diophantine core of the book in Parts C, D, and E. The first of these, Part C, contains a proof of the Mordell–Weil theorem: *The group of rational points on an abelian variety is finitely generated.* It also includes in Sections C.4 and C.5 a discussion of Galois cohomology and the Selmer and Tate–Shafarevich groups, which are used for studying the refined properties of the Mordell–Weil group. Next, in Part D, we give a proof of Roth's theorem: *There are only finitely many rational numbers that approximate a given algebraic number to order* $2+\varepsilon$. We then use this fundamental theorem on Diophantine approximation and the arithmetic–geometric relations provided by the theory of heights to prove Siegel's theorem: *A curve of genus* $g \geq 1$ *has only finitely many integer points.* Finally, in Part E we take up the question of curves of higher genus and prove Mordell's 1922 conjecture (Faltings [1], 1983): *A curve of genus* $g \geq 2$ *has only finitely many rational points.* The proof that we give is based on Diophantine approximation techniques similar to those used in the proof of Roth's theorem. This alternative proof of Faltings' theorem is due to Vojta [1], with substantial simplifications by Bombieri [1].

The preceding material easily fills the present volume, but leaves unmentioned many important Diophantine results and an even larger number of important Diophantine conjectures. As a means of introducing the reader

to this additional material, we include in Part F an overview of further results and open problems. Topics covered include rational and algebraic points on curves and abelian varieties, the discreteness of algebraic points relative to the height metric, bounds for height functions (both proven and conjectural), the search for effectivity in the Mordell–Weil theorem and in Faltings' theorem, and a further discussion of how geometry governs arithmetic, including deep conjectures of Batyrev, Bombieri, Lang, Manin, and Vojta that provide much of the focus for current research in Diophantine geometry.

Readers should be aware that even with the survey material included in Part F, we have been forced to leave out or only touch upon many topics that are relevant to the Diophantine problems studied in this volume. These topics include:

(a) *Baker's Method*
Effectivity in Diophantine geometry is discussed in Part F, but at present the only general effective theorems come from Baker's method giving lower bounds for linear forms in logarithms. Since it would be impossible to do justice to this vast subject without significantly increasing the size of the present volume, we content ourselves with quoting an exemplary result in (D.9.5).

(b) *Arakelov Geometry*
During the past fifteen years, Arakelov geometry has been one of the main sources of inspiration both for developing the theory of Diophantine geometry and for solving Diophantine problems. We give a motivated introduction in Section B.10, but again even a complete volume (such as Lang [7]) is hardly enough to do justice to the subject.

(c) *Existence of Rational Points*
Most of the principal theorems in this volume assert the finiteness of the set of rational points on certain varieties or, failing that, give an estimate for the number of rational points of bounded height. We thus do not address the important problem of deciding whether a variety possesses any rational points at all. The main tools for addressing this important Diophantine problem are cohomological. We discuss this question in Part C, but only for homogeneous spaces of abelian varieties as it relates to the Mordell–Weil theorem.

(d) *Function Fields*
The celebrated analogies between number fields and function fields are discussed in Section A.9. These form the starting point of Arakelov theory. Indeed, the theory of heights and all of the main theorems proven in this volume can be described in a common language over both number fields and function fields, or more generally over finitely generated fields, as is done in the seminal work of Lang [6]. We apologize for our lack of generality, but we note that there are often better methods involving the use of derivations available in the function field case that are unavailable when one is working over number fields.

Prerequisites

The main prerequisite for reading this book is a solid understanding of basic algebraic number theory, including such topics as rings of integers, completions, ramification, ideal class groups, and unit groups. This material is covered in any standard text such as Lang [9]. A second prerequisite for understanding the main theorems in this book is a working knowledge of algebraic geometry. In order to make this volume as self-contained as possible, we have included an introduction to algebraic geometry in Part A; but it is also a truism that when studying Diophantine problems, one can never know too much algebraic geometry, so any previous exposure is sure to be helpful.

References and Exercises

We have divided the book into six lettered parts, A–F, and each part is divided into sections and subsections. Items in each section are numbered consecutively, and cross-references are given in full, for example (A.8.2.2) or (E.10.3). Exercises appear at the end of each part, except for the lengthy part A, which has exercises at the end of each section. Exercises are numbered consecutively and are fully referenced, so for example, Exercise A.4.6 is the sixth exercise in Section A.4, and Exercise E.5 is the fifth exercise in Part E. Bibliographic references are given by the author's name followed by a reference number in square brackets, for example Tate [3, Theorem 2].

This volume contains numerous exercises. The reader desiring to gain a real understanding of the subject is urged to attempt as many as possible. Some of these exercises are (special cases of) results that have appeared in the literature. A list of comments and citations for the exercises will be found at the end of the book. Exercises marked with a single asterisk are somewhat more difficult, and two asterisks signal an unsolved problem.

Standard Notation

Throughout this book, we use the symbols

$$\mathbb{Z},\ \mathbb{Q},\ \mathbb{R},\ \mathbb{C},\ \mathbb{F}_q,\ \text{and } \mathbb{Z}_p$$

to represent the integers, rational numbers, real numbers, complex numbers, field with q elements, and p-adic integers, respectively. Further, if R is any ring, then R^* denotes the group of invertible elements of R; and if A is an abelian group, then A_m or $A[m]$ denotes the subgroup of A consisting of all elements whose order divides m. A detailed list of notation will be found at the end of the book.

PART A

The Geometry of Curves and Abelian Varieties

The heavens rejoice in motion, why should I
Abjure my so much lov'd variety.
John Donne, *Elegies*

Do Not Read
Part A

Now that we have your attention, let us explain why we would recommend that you *not* read part of this book. Part A contains a summary of the main results from algebraic geometry that will be needed in our arithmetic investigations. If you begin your study of Diophantine geometry by attempting to read all of Part A and doing all of the exercises, you are likely to feel overwhelmed by the geometry before you reach any of the beautiful arithmetic results. So we suggest that you begin by skimming Part A, possibly reading more closely any material that covers gaps in your knowledge. Then as you read the rest of this book, use Part A as a reference source for geometric facts as they are needed. Having offered this warning and advice, we now begin our (far from brief) survey of algebraic geometry.

A general principle suggests that before tackling a Diophantine problem, it is necessary first to understand the underlying geometry. The initial part of this book develops the geometry necessary to do arithmetic on curves, that is, on algebraic varieties of dimension one. However, we cannot be content to work only with varieties of dimension one. For example, we will want to work with surfaces that are the product of two curves. More importantly, many of the deeper properties of a curve are best analyzed by studying a certain variety of higher dimension called the Jacobian of the curve. The Jacobian of a curve is a group variety. It represents a kind of linearization of the curve, that is, it is a space where we can add points on the curve to one another. Jacobians are special instances of abelian varieties, and the theory gains unity when developed in this generality.

After a brief survey of the basic concepts of algebraic geometry in a preliminary section, we describe with a bit more detail divisors and linear systems on varieties in the next two sections. In Section A.4 we give a succinct account of the geometry of curves, centered on the notion of genus

and the Riemann–Roch theorem. We then pursue the theory of Jacobians and abelian varieties in the next four sections. These four sections form the core of this geometric part. We develop first the theory of abelian varieties and Jacobians analytically over $\mathbb{C}$ and then algebraically over an arbitrary field, often giving two proofs or at least offering two different perspectives on parallel results.

Ultimately, we want to study arithmetic, that is, points defined over number fields. So the reader may be surprised that we devote so much space to complex varieties. There are several reasons for this. The first is historical. Algebraic geometry was originally developed (by Riemann and others) as a part of complex function theory, and imitating history often gives valuable insight into a subject. Secondly, there is the so-called Lefschetz principle, which says that geometry over any algebraically closed field of characteristic 0 is essentially the same as over $\mathbb{C}$. This metaprinciple is true because any field of characteristic 0 and finite (or countable) transcendence degree over $\mathbb{Q}$ can be embedded into $\mathbb{C}$, and virtually all objects of an algebraic nature are defined over such fields. Further, Galois theory provides a tool that often allows one to descend from the algebraic closure of $\mathbb{Q}$ to a number field (see Proposition A.2.2.10 for a precise statement). A third, equally compelling, reason to study complex varieties is the philosophy of Arakelov, which casts the complex points of a variety as the "fiber at infinity" that "compactifies" a model of the variety over $\mathbb{Z}$. We will be able to give only a brief introduction to these ideas and will not rely on them for proofs, but their importance in the development of Diophantine geometry (past and future) can hardly be overstated. Arakelov's philosophy was utilized by Faltings in the original proof of Mordell's conjecture, and the generalization of Arakelov's ideas to higher dimensions played a very significant role in the second proof, found by Vojta.

However, working over $\mathbb{C}$ is clearly not sufficient for our needs. It is important to know that constructions such as the Jacobian of a curve can be done over the field of definition of the curve. Also, when studying varieties defined over a number field, one is naturally led to specialize them "mod p." This requires geometry in characteristic p. So we will also need to employ the tools of abstract algebraic geometry.

It will not be possible for us to provide full proofs of all of the statements in this part. Instead, we will state general theorems and definitions of algebraic geometry and provide adequate references. We give more details on the specific applications to curves and abelian varieties, but even here we have had to omit some important results due to lack of space and time. We have tried to keep our baggage to a minimum, often at the risk of appearing "old-fashioned."

Finally, a word of caution. Although the geometry we develop will suffice to prove the Mordell–Weil theorem and Faltings' theorem (Mordell's conjecture), there is little doubt that further progress is likely to require the sophisticated apparatus of modern algebraic geometry. The language

of schemes is essential both for its powerful technical versatility and for the valuable insights it provides for arithmetic geometry. We give an introduction to this deep subject in a last section presented as a variation on the old theme of "analogies between function fields and number fields." Especially, we describe what is meant by a curve "over $\mathbb{Z}$" and by an abelian variety "over $\mathbb{Z}$," and we explain what "reducing modulo p" means in this context.

General references on algebraic geometry include Hartshorne [1], Griffiths–Harris [1], Shafarevich [1], and Mumford [4, 6]. There is a vast literature on curves; we mention Walker [1], Fulton [1], and at the other end of the spectrum, Arbarelo–Cornalba–Griffiths–Harris [1]. Textbooks on Jacobians and abelian varieties are rarer. Complex abelian varieties and theta functions are nicely introduced in Swinnerton-Dyer [1], K. Murty [1], and Lang [4], and thoroughly treated in Lange and Birkenhake [1]. The algebraic aspects may be found in Lang [3], Mumford [2], and Weil's original books [2, 3]. The survey of Bost [1] provides an excellent presentation of Jacobian varieties, and Mumford's lectures [3] give a pleasant account of curves and their Jacobians. Jacobians are also treated in an analytical fashion in Griffiths–Harris [1] and Gunning [1], while the algebraic construction of Weil is described in Serre [1]. Finally, we point out the excellent surveys and bibliographies in the papers of Rosen [1] and Milne [1, 2].

A.1. Algebraic Varieties

This preliminary chapter is essentially a glossary and a herbarium. We review the basic definitions of algebraic geometry and collect examples of varieties and maps. It can safely be omitted by any reader with some knowledge of algebraic geometry. Throughout we work with the following notation:

k	a perfect field.
$\bar{k}$	an algebraic closure of k.
G_k	$= \mathrm{Gal}(\bar{k}/k)$, the Galois group of $\bar{k}$ over k.

The reason for working in this generality is that we want to be able to study fields k of arithmetic interest, such as $\mathbb{Q}$, $\mathbb{Q}_p$, or $\mathbb{F}_p$, but geometric properties are best expressed over algebraically closed fields. As a naive example, we might consider the equation $x^2 + y^2 + 1 = 0$ as giving a curve defined over $\mathbb{Q}$, yet this curve is an empty set in the sense that it has no points with x and y in $\mathbb{Q}$. Hence to "see" the curve, we must look at all the points with coordinates in $\bar{\mathbb{Q}}$. The restriction to perfect fields is usually not essential, but it is made to simplify our work. Especially, the notion of "being defined over k" is unambiguous in this context (see Exercise A.1.13).

A.1.1. Affine and Projective Varieties

We begin our review with affine n-space.

Definition. *Affine n-space* (over k), which we denote by $\mathbb{A}^n$ or $\mathbb{A}^n_k$, is the set

$$\mathbb{A}^n = \{(x_1, \dots, x_n) \mid x_i \in \bar{k}\}.$$

The *set of k-rational points of* $\mathbb{A}^n$ is the set

$$\mathbb{A}^n(k) = \{(x_1, \dots, x_n) \in \mathbb{A}^n \mid x_i \in k\}.$$

Remarks. (i) One may also characterize the set of k-rational points of $\mathbb{A}^n$ as the set

$$\mathbb{A}^n(k) = \{(x_1, \dots, x_n) \in \mathbb{A}^n \mid \sigma(x_i) = x_i \text{ for all } \sigma \in G_k\}.$$

(ii) The notation $\mathbb{A}^n(k)$ is, in fact, that of a functor. The functor $\mathbb{A}^n$ associates to each field k the set $\mathbb{A}^n(k)$.

Now let I be an ideal in $\bar{k}[X_1, \dots, X_n] = \bar{k}[X]$. We associate to I its set of zeros,

$$Z(I) = \{x \in \mathbb{A}^n \mid P(x) = 0 \text{ for all } P \in I\}.$$

In some sense the primary goal of algebraic geometry is to understand the spaces thus defined. Similarly, to each subset S of $\mathbb{A}^n$ we associate the ideal of polynomials vanishing on S,

$$I_S = \{P \in \bar{k}[X] \bigm| P(x) = 0 \text{ for all } x \in S\}.$$

Definition. An *affine algebraic set* S is a set of the form $S = Z(I)$ for some ideal I in $\bar{k}[X]$. The set S is said to be *defined over* k if its ideal I_S can be generated by polynomials in $k[X]$.

For example, a point $a = (a_1, \dots, a_n)$ is an algebraic set defined by the ideal generated by the polynomials $x_1 - a_1, \dots, x_n - a_n$, and obviously it is defined over k if and only if each a_i belongs to k.

Remarks. (i) The Hilbert basis theorem says that any ideal of polynomials is generated by a finite number of polynomials. See, for example, Atiyah–Macdonald [1, Theorem 7.5] or Lang [2, Section 6.2]. Thus algebraic sets can always be written as the common zeros of a finite collection of polynomials.

(ii) If V is an algebraic set defined over k by some ideal I, then its set of k-rational points is defined by

$$\begin{aligned} V(k) &= \{x \in \mathbb{A}^n(k) \mid P(x) = 0 \text{ for all } P \in I\} \\ &= \{x \in V \mid \sigma(x) = x \text{ for all } \sigma \in G_k\}. \end{aligned}$$

(iii) It is also convenient to define

$$I_{V,k} = \{P \in k[X] \mid P(x) = 0 \text{ for all } x \in V\}.$$

Note that we always have $I_{V,k} \cdot \bar{k}[X] \subset I_{V,\bar{k}} = I_V$, and equality occurs exactly when V is defined over k.

Recall that for any ring R, the *radical* $\sqrt{I}$ of an ideal $I \subset R$ is defined to be

$$\sqrt{I} = \{a \in R \mid a^r \in I \text{ for some } r \geq 1\}.$$

We now describe the correspondence between algebraic sets and polynomial ideals.

Lemma A.1.1.1. (i) *Let V_i be algebraic subsets of $\mathbb{A}^n$. Then arbitrary intersections $\bigcap_i V_i$ and finite unions $V_1 \cup \cdots \cup V_r$ are algebraic sets.*

(ii) *If $S_1 \subset S_2 \subset \mathbb{A}^n$, then $I_{S_1} \supset I_{S_2}$.*

(iii) *If $I_1 \subset I_2 \subset \bar{k}[X]$, then $Z(I_1) \supset Z(I_2)$.*

(iv) *If V is an algebraic set, then $Z(I_V) = V$.*

(v) *If I is an ideal in $\bar{k}[X]$, then $I_{Z(I)} = \sqrt{I}$.*

PROOF. Statement (i) is clear from the equalities $\bigcap_i V_i = Z(\sum_i I_{V_i})$ and $V_1 \cup V_2 = Z(I_{V_1} \cdot I_{V_2})$. The rest is easy except for (v), which is a consequence of the next theorem. □

Theorem A.1.1.2. (Hilbert's Nullstellensatz) *Let I be an ideal of the ring $\bar{k}[X_1, \ldots, X_n]$ and let P be a polynomial vanishing at every point in $Z(I)$. Then there is an integer $r \geq 1$ such that $P^r \in I$.*

PROOF. See Lang [2, Section 10.2] or Atiyah–Macdonald [1, Chapter 7, Exercise 14]. It is, of course, essential to formulate the theorem over $\bar{k}$. □

Lemma A.1.1.1 says that there is a natural bijection between algebraic sets and reduced ideals, that is, ideals that are equal to their own radical. The first part of Lemma A.1.1.1 can be reformulated by saying that algebraic sets satisfy the axioms of the closed sets of a topology. Note that $\mathbb{A}^n$ and the empty set are algebraic sets, since

$$\mathbb{A}^n = Z(\{0\}) \qquad \text{and} \qquad \emptyset = Z(\bar{k}[X]).$$

Definition. The *Zariski topology* on $\mathbb{A}^n$ is the topology whose closed sets are algebraic sets. The *Zariski topology* on an algebraic set S is the topology induced by the inclusion $S \subset \mathbb{A}^n$.

Definition. A nonempty subset Z of a topological space X is *irreducible* if it cannot be written as the union of two proper closed subsets of Z (for the induced topology).

Example. $\mathbb{A}^n$ is irreducible for the Zariski topology. To see this, we observe that the Zariski topology is highly non-Hausdorff. Indeed, any nonempty open subset of $\mathbb{A}^n$ is dense in $\mathbb{A}^n$, and hence the intersection of any two such open sets is always nonempty.

Definition. An *affine variety* is an irreducible algebraic subset (for the Zariski topology) of some $\mathbb{A}^n$.

Lemma A.1.1.3. (i) *An algebraic set V is irreducible if and only if its ideal I_V is a prime ideal.*

(ii) *An algebraic set is a finite union of varieties. If we insist, as we may, that none of the varieties be contained in another one, then this decomposition is unique.*

The varieties in the decomposition (ii) of an algebraic set are called the *irreducible components* of the algebraic set.

PROOF. Easy. See Hartshorne [1, I.1.6]. □

Example A.1.1.4. (Affine hypersurfaces) Let $P \in \bar{k}[X_1, \ldots, X_n]$ be a polynomial and let $V = Z(P)$ be the algebraic set defined by P. Suppose that $P = P_1^{m_1} \cdots P_r^{m_r}$ is the decomposition of P into irreducible factors, and set $V_i = Z(P_i)$. Then the V_i's are the irreducible components of V. Indeed, each V_i is a variety, $V = \bigcup_{i=1}^r V_i$, and $V_i \not\subset V_j$ for $i \neq j$.

The algebra of polynomials in n variables is naturally associated to the affine space $\mathbb{A}^n$. When we restrict polynomial functions to an affine subvariety V, it is natural to identify any two polynomials that give the same function on V. Thus we are led to the following definition:

Definition. Let V be an affine subvariety of $\mathbb{A}^n$. The *affine coordinate ring of V* is

$$\bar{k}[V] = \bar{k}[x_1, \ldots, x_n]/I_V.$$

We will see in the next section that this algebra completely characterizes the variety V.

Example A.1.1.5. (Products of affine varieties) We observe that there is an obvious isomorphism $\mathbb{A}^m \times \mathbb{A}^n \cong \mathbb{A}^{m+n}$ given by the map

$$((x_1, \ldots, x_m), (y_1, \ldots, y_n)) \longmapsto (x_1, \ldots, x_m, y_1, \ldots, y_n).$$

(Although "isomorphism" is not formally defined until the next section, the meaning is clear here.) If $V \hookrightarrow \mathbb{A}^m$ and $W \hookrightarrow \mathbb{A}^n$ are two affine varieties, then we define their product $V \times W$ to be the affine variety whose ideal is generated by I_V and I_W inside $\bar{k}[x_1, \ldots, x_m, y_1, \ldots, y_n]$. It is not hard to verify that

$$\bar{k}[V \times W] \cong \bar{k}[V] \otimes \bar{k}[W].$$

(See Hartshorne [1, I, Exercise 3.15]. This is still true with k in place of $\bar{k}$, provided that we keep the assumption that k is perfect.)

Since at least the work of Desargues it has been known that geometry is easier if one adds "points at infinity" in order to make affine space "complete." For example, one wants the following kinds of statements to be true: Two distinct lines in the plane meet in one point, a line meets a conic in two points (counted with multiplicities), etc. Clearly, these statements are false in the affine plane $\mathbb{A}^2$, since parallel lines do not meet. In order to make them true, we introduce projective space.

Definition. *Projective n-space* $\mathbb{P}^n$ is the set of lines through the origin in $\mathbb{A}^{n+1}$. In symbols,

$$\mathbb{P}^n = \frac{\{(x_0,\ldots,x_n) \in \mathbb{A}^{n+1} \mid \text{some } x_i \neq 0\}}{\sim} = \frac{\mathbb{A}^{n+1} \smallsetminus \{0\}}{\sim},$$

where the equivalence relation $\sim$ is defined by

$$(x_0,\ldots,x_n) \sim (y_0,\ldots,y_n) \iff (x_0,\ldots,x_n) = \lambda(y_0,\ldots,y_n) \text{ for some } \lambda \in \bar{k}^*.$$

If $P \in \mathbb{P}^n$ is the point representing the equivalence class of the $(n+1)$-tuple $(x_0,\ldots,x_n)$, the x_i's are called *homogeneous* or *projective coordinates* for the point P. The *set of k-rational points of* $\mathbb{P}^n$, denoted by $\mathbb{P}^n(k)$, is the set of lines through the origin in $\mathbb{A}^{n+1}$ that are defined over k. This is the set of points in $\mathbb{P}^n$ for which we can find some homogeneous coordinates in $\mathbb{A}^{n+1}(k)$. Equivalently, these are the points $(x_0,\ldots,x_n)$ with the property that for any nonzero coordinate x_j, all of the ratios x_i/x_j are in k.

The Galois group G_k acts on $\mathbb{P}^n$ by acting on the coordinates,

$$\sigma(P) = \bigl(\sigma(x_0),\ldots,\sigma(x_n)\bigr) \qquad \text{for } P = (x_0,\ldots,x_n) \in \mathbb{P}^n \text{ and } \sigma \in G_k.$$

Then one can show that (Exercise A.1.16)

$$\mathbb{P}^n(k) = \bigl\{P \in \mathbb{P}^n \mid \sigma(P) = P \text{ for all } \sigma \in G_k\bigr\}.$$

The *field of definition* of a point $P = (x_0,\ldots,x_n) \in \mathbb{P}^n$ is the smallest extension of k over which P is rational, namely,

$$k(P) := k(x_0/x_j, x_1/x_j, \ldots, x_n/x_j) \qquad \text{for any } j \text{ with } x_j \neq 0.$$

Equivalently, $k(P)$ is determined by the property

$$\operatorname{Gal}\bigl(\bar{k}/k(P)\bigr) = \{\sigma \in G_k \mid \sigma(P) = P\}.$$

In order to define projective algebraic sets, we recall that a polynomial ideal is homogeneous if it is generated by homogeneous polynomials, or, alternatively, if the homogeneous components of any polynomial in the ideal are again in the ideal. If P is a homogeneous polynomial, then

$$P(x_0,\ldots,x_n) = 0 \iff P(\lambda x_0,\ldots,\lambda x_n) = 0 \text{ for all } \lambda \in \bar{k}^*.$$

We can thus define projective algebraic sets in a fashion entirely analogous to our definition of affine algebraic sets, provided that we use homogeneous polynomials and ideals.

Definition. A *projective algebraic set* is the set of zeros in $\mathbb{P}^n$ of a homogeneous ideal in $\bar{k}[x_0, \ldots, x_n]$. The *Zariski topology on* $\mathbb{P}^n$ is defined by taking the projective algebraic sets to be the closed sets, and the Zariski topology on an algebraic set is the topology induced from the Zariski topology on $\mathbb{P}^n$. A *projective variety* is a projective irreducible algebraic set. It is said to be *defined over* k if its ideal can be generated by polynomials in $k[x_0, \ldots, x_n]$.

The correspondence between homogeneous ideals and projective algebraic sets is very similar to the affine one; the only difference is the existence of an "irrelevant" ideal, namely the ideal I_0 generated by $x_0, \ldots, x_n$. Notice that I_0 defines the empty subset of $\mathbb{P}^n$, and any homogeneous ideal different from $\bar{k}[x_0, \ldots, x_n]$ is contained in I_0. Let us define a *saturated* ideal as a homogeneous ideal I such that if $x_i f \in I$ for all $i = 0, \ldots, n$, then $f \in I$; clearly, the ideal of polynomials vanishing on a projective algebraic set is saturated. More precisely, the map $I \mapsto Z(I)$ gives a bijection between reduced saturated ideals and projective algebraic sets. Further, a projective algebraic set Z is a projective variety if and only if I_Z is a (homogeneous) prime ideal in $\bar{k}[x_0, \ldots, x_n]$.

Example A.1.1.6. A variety defined by linear forms

$$L_1(x_0, \ldots, x_n) = \cdots = L_r(x_0, \ldots, x_n) = 0$$

is called a *linear subvariety* of $\mathbb{P}^n$. For example, a point with projective coordinates $(a_0, \ldots, a_n)$ is defined by the linear forms $a_i x_j - a_j x_i = 0$. An algebraic set defined by one nonzero homogeneous polynomial is called a *projective hypersurface.* A linear hypersurface is called a *hyperplane.*

Just as with affine varieties, we look at the quotient of the polynomial algebra by the homogeneous ideal of a projective variety.

Definition. The *homogeneous coordinate ring* of a projective variety $V \subset \mathbb{P}^n$ is the quotient

$$S(V) = \bar{k}[x_0, \ldots, x_n]/I_V.$$

Note that unlike the case of $k[V]$ for affine varieties, the elements of $S(V)$ do not define functions on a projective variety V. An even more important observation is that the homogeneous coordinate ring depends on the embedding of V in $\mathbb{P}^n$, it is not an intrinsic invariant of V (see Exercise A.1.4).

Let us explain now how to cover $\mathbb{P}^n$ (or any projective variety) by affine spaces and thereby recover the classical description of $\mathbb{P}^n$ as the union of affine space $\mathbb{A}^n$ together with a hyperplane at infinity.

Definition. Let $(x_0, \ldots, x_n)$ be homogeneous coordinates on $\mathbb{P}^n$. The *standard (affine) open subset* U_i is the complement of the hyperplane defined by $x_i = 0$.

It is obvious that the open sets U_i cover $\mathbb{P}^n$, and it is easy to see that the map

$$\mathbb{A}^n \longrightarrow U_i, \qquad (a_1, \dots, a_n) \longmapsto (a_1, \dots, a_{i-1}, 1, a_{i+1}, \dots, a_n),$$

is a homeomorphism with inverse

$$(x_0, \dots, x_n) \longmapsto (x_0/x_i, \dots, x_{i-1}/x_i, x_{i+1}/x_i, \dots, x_n/x_i).$$

Thus U_i is isomorphic to $\mathbb{A}^n$, and since the hyperplane $x_i = 0$ is isomorphic to $\mathbb{P}^{n-1}$, we obtain a description of $\mathbb{P}^n$ as the union of affine space with a hyperplane at infinity. Repeating this process gives a cellular decomposition

$$\mathbb{P}^n = \mathbb{A}^n \cup \mathbb{P}^{n-1} = \cdots = \mathbb{A}^n \cup \mathbb{A}^{n-1} \cup \cdots \cup \mathbb{A}^1 \cup \mathbb{A}^0.$$

Example A.1.1.7. The completion of an affine variety to a projective variety can be done very concretely by homogenizing the polynomials defining it. Similarly, one can find an open affine subset of a projective variety by dehomogenizing its defining polynomials.

For example, let U be the affine parabola defined by $y - x^2 = 0$ in $\mathbb{A}^2$. Then the homogeneous equation $ZY - X^2 = 0$ defines a projective variety V, and the map $(X, Y, Z) \mapsto (X/Z, Y/Z)$ defines an isomorphism from $V \cap \{Z \neq 0\}$ to U. Similarly, the set $V \cap \{X \neq 0\}$ is isomorphic to the affine hyperbola $uv - 1 = 0$ by the map $(X, Y, Z) \mapsto (Y/X, Z/X)$. Notice that the parabola has one point at infinity, while the hyperbola has two.

It is convenient to be able to speak of open subsets of varieties as varieties themselves, so we enlarge our category a bit.

Definition. A *quasi-projective algebraic set* is an open subset of a projective algebraic set. A *quasi-projective variety* is an irreducible quasi-projective algebraic set.

Notice that affine and projective varieties are quasi-projective, but there are quasi-projective varieties that are neither affine nor projective. For example, $\mathbb{P}^2 \setminus \{(0,0,1)\}$ is quasi-projective, but it is neither affine nor projective. On the other hand, any quasi-projective variety can be covered by affine open subsets, because the complement of a hypersurface in $\mathbb{A}^n$ is an affine variety. (See Hartshorne [1, I.4.2 and Exercises I.3.5, I.3.6]). This suggests the following principle: Global properties are better studied in the context of projective varieties, whereas local properties are most easily verified on open affine sets.

A.1.2. Algebraic Maps and Local Rings

Having defined algebraic varieties, we need to define maps between them. For many reasons, we want to have a coordinate-free approach and to consider varieties independent of any particular embedding in $\mathbb{A}^n$ or $\mathbb{P}^n$. Roughly speaking, an algebraic map between varieties is a map that can be defined by polynomials or rational functions. We start by defining the functions on a variety X, that is, maps from X to $\mathbb{A}^1 = \bar{k}$.

Definition. Let X be a variety and x' a point on X. A function $f : X \to \bar{k}$ is *regular at* x' if there exists an open affine neighborhood $U \subset X$ of x', say $U \subset \mathbb{A}^n$, and two polynomials $P, Q \in \bar{k}[x_1, \ldots, x_n]$ such that $Q(x') \neq 0$ and $f(x) = P(x)/Q(x)$ for all $x \in U$. The function f is *regular on* X if it is regular at every point of X. The *ring of regular functions on* X is denoted by $\mathcal{O}(X)$.

Note that if f is regular on X, it need not be true that there are fixed polynomials P, Q such that $f = P/Q$ at every point of X, although this will be true for affine varieties (see Theorem A.1.2.1 below). The definition of regularity is local, so it may be necessary to choose different polynomials at different points. More precisely, if f is regular on X, then one can write X as a finite union of affine open subsets U_i, and one can find polynomials P_i, Q_i such that $f(x) = P_i(x)/Q_i(x)$ for all $x \in U_i$.

We also note that the property of being regular is open. If f is regular at x, then it is regular at every point in some neighborhood of x. This suggests looking at the collection of functions that are regular at a given point.

Definition. Let x be a point on a variety X. The *local ring of* X *at* x is the ring of functions that are regular at x, where we identify two such functions if they coincide on some open neighborhood of x. This ring is denoted by $\mathcal{O}_{x,X}$, or simply by $\mathcal{O}_x$ if no confusion is likely to arise.

More generally, we can define the ring of functions regular along a subvariety of X.

Definition. Let X be a variety and $Y \subset X$ a subvariety. The *local ring of* X *along* Y, denoted by $\mathcal{O}_{Y,X}$, is the set of pairs (U, f), where U is an open subset of X with $U \cap Y \neq \emptyset$ and $f \in \mathcal{O}(U)$ is a regular function on U, and where we identify two pairs $(U_1, f_1) = (U_2, f_2)$ if $f_1 = f_2$ on $U_1 \cap U_2$. The ring $\mathcal{O}_{Y,X}$ is a local ring, its unique maximal ideal being given by

$$\mathcal{M}_{Y,X} = \{f \in \mathcal{O}_{Y,X} \mid f(x) = 0 \text{ for all } x \in Y\}.$$

For example, $\mathcal{O}_{\{x\},X}$ is just the local ring of X at x, while the local ring $\mathcal{O}_{X,X}$ turns out to be a field.

Definition. Let X be a variety. The *function field of* X, denoted by $\bar{k}(X)$, is defined to be $\mathcal{O}_{X,X}$, the local ring of X along X. In other words, $\bar{k}(X)$ is the set of pairs (U, f), where U is an open subset of X and f is a regular function on U, subject to the identification $(U_1, f_1) = (U_2, f_2)$ if $f_1 = f_2$ on $U_1 \cap U_2$. (N.B. An element f of $\bar{k}(X)$ is not a function defined at every point of X. Instead, f is a function that is defined at some point of X, and hence is defined on a nonempty open set of points of X.)

It is easy to check that $\bar{k}(X)$ is a field that contains every local ring $\mathcal{O}_{Y,X}$ of X, and that for any subvariety $Y \subset X$, we have $\mathcal{O}_{Y,X}/\mathcal{M}_{Y,X} \cong \bar{k}(Y)$. The function fields of $\mathbb{A}^n$ and $\mathbb{P}^n$ are both equal to $\bar{k}(x_1, \dots, x_n)$, the field of rational functions in n indeterminates. If X is an affine hypersurface defined by an irreducible polynomial $P(x_1, \dots, x_n)$ in which the variable x_n appears, then $\bar{k}(X)$ is an algebraic extension of $\bar{k}(x_1, \dots, x_{n-1})$ generated by any root α of the equation $P(x_1, \dots, x_{n-1}, \alpha) = 0$. The local ring of a point $(a_1, \dots, a_n) \in \mathbb{A}^n$ is the ring of polynomials $\bar{k}[x_1, \dots, x_n]$ localized at the ideal $(x_1 - a_1, \dots, x_n - a_n)$. We are now ready to define maps between varieties.

Definition. A map $\phi : X \to Y$ between varieties is a *morphism* if it is continuous, and if for every open set $U \subset Y$ and every regular function f on U, the function $f \circ \phi$ is regular on $\phi^{-1}(U)$. A map is *regular at a point* x if it is a morphism on some open neighborhood of x.

In a less intrinsic way, one can show that f is regular at x if there is an affine neighborhood $U \subset \mathbb{A}^m$ of x in X and an affine neighborhood $V \subset \mathbb{A}^n$ of $\phi(x)$ in Y such that ϕ sends U into V and such that ϕ can be defined on U by n polynomials in m variables. That these definitions are equivalent comes from the fact that a morphism of affine varieties is defined globally by polynomials, as can be deduced readily from Theorem A.1.2.1 below. The word "morphism" is short for "morphism in the category of algebraic varieties." Just as with rational functions, it is often convenient to consider maps between varieties that are defined only on an open subset. We therefore introduce one more definition.

Definition. A *rational map* from a variety X to a variety Y is a map that is a morphism on some nonempty open subset of X. A rational map $\phi : X \to Y$ is said to be *dominant* if $\phi(U)$ is dense in Y for some (and consequently every) nonempty open set $U \subset X$ on which it is a morphism.

A *birational map* is a rational map that has a rational inverse. Two varieties are said to be *birationally equivalent* if there is a birational map between them.

Remark. Let $\phi : X \to Y$ be a rational map. Then there is a largest open subset U on which ϕ is a morphism. This open subset is called the *domain of* ϕ.

We have defined morphisms and rational maps purely in terms of local properties. We now examine their global behavior, distinguishing carefully

between affine and projective varieties. Notice that almost by definition, a morphism $\phi : V \to W$ of affine varieties induces a ring homomorphism $\phi^* : \bar{k}[W] \to \bar{k}[V]$ defined by $f \mapsto f \circ \phi$.

Theorem A.1.2.1. (i) *Let V be an affine variety. Then $\mathcal{O}(V) \cong \bar{k}[V]$.*
(ii) *Let V, W be affine varieties. The natural map*

$$\begin{aligned} \mathrm{Mor}(V,W) &\longrightarrow \mathrm{Hom}_{\bar{k}\text{-Alg}}(\bar{k}[W], \bar{k}[V]), \\ \phi &\longmapsto (f \mapsto f \circ \phi), \end{aligned}$$

is a bijection. In fancy language, the association $V \to \bar{k}[V]$ is a contravariant functor that induces an equivalence between the category of affine varieties and the category of finitely generated integral $\bar{k}$-algebras.

PROOF. Hartshorne [1, I.3.2]. □

Thus an affine variety is completely determined by its ring of regular functions. This stands in stark contrast to the next two results.

Lemma A.1.2.2. *A regular function on a projective variety is constant.*

PROOF. Hartshorne [1, I.3.4(a)]. □

Theorem A.1.2.3. *The image of a projective variety by a morphism is a projective variety. More generally, if X is a projective variety, the projection $X \times Y \to Y$ is a closed map.*

PROOF. This is essentially equivalent to the main theorem of elimination theory; see Van der Waerden [1, vol. II, §80] or Shafarevich [1, I.2 Theorems 2, 3].

□

Notice that the image of an affine variety by a morphism need not be an affine variety, so there is no analogue of Theorem A.1.2.3 for affine varieties.

We now look at local rings and function fields. Recall that if $\mathfrak{p}$ is a prime ideal in a ring A, then the localized ring at $\mathfrak{p}$ is

$$A_{\mathfrak{p}} = \left\{ \frac{a}{b} \;\middle|\; a, b \in A,\ b \notin \mathfrak{p} \right\}.$$

If $\mathfrak{p}$ is a homogeneous ideal in a graded ring A, the homogeneous localized ring at $\mathfrak{p}$ is

$$A_{(\mathfrak{p})} = \left\{ \frac{a}{b} \;\middle|\; a, b \in A,\ \deg(a) = \deg(b),\ b \notin \mathfrak{p} \right\}.$$

In both cases, the local ring is a subring of the ring of fractions of A, which we denote by $\mathrm{Frac}(A)$. Of course, if A is a domain, then $\mathrm{Frac}(A)$ is a field. For the general theory of localization, see, for example, Lang [2, II.3] or Matsumura [1, I.1].

Theorem A.1.2.4. (i) *Let P be a point on an affine variety V, and let $\mathcal{M}_P$ be the ideal of functions in $\bar{k}[V]$ that vanish at P. Then*

$$\mathcal{O}_{P,V} = \bar{k}[V]_{\mathcal{M}_P} \qquad \text{and} \qquad \bar{k}(V) = \operatorname{Frac}(\bar{k}[V]).$$

(ii) *Let P be a point on a projective variety V, and let $\mathcal{M}_P$ be the ideal generated by homogeneous polynomials vanishing at P. Then*

$$\mathcal{O}_{P,V} = S[V]_{(\mathcal{M}_P)} \qquad \text{and} \qquad \bar{k}(V) = S(V)_{((0))}.$$

PROOF. (i) Hartshorne [1, I.3.2(c,d)].
(ii) Hartshorne [1, I.3.4(b,c)]. □

Notice that the elements of $S[V]_{(\mathcal{M}_P)}$ may be viewed as functions on V because they are of degree zero, which means that their value at a point is independent of the choice of homogeneous coordinates for that point. Of course, they need not be defined at every point of V.

Theorem A.1.2.5. *Let $f : V \to W$ be a rational map between two varieties.*

(i) *If f is regular at P and $Q = f(P)$, then the map*

$$f^* : \mathcal{O}_{Q,W} \longrightarrow \mathcal{O}_{P,V}, \qquad f^* : g \longmapsto g \circ f,$$

is a homomorphism of local rings. In particular, $f^*(\mathcal{M}_Q) \subset \mathcal{M}_P$.

(ii) *If f is dominant, then f^* defines a field homomorphism $\bar{k}(W) \hookrightarrow \bar{k}(V)$. Conversely, every such field homomorphism corresponds to a dominant rational map. In other words, the association $X \to \bar{k}(X)$ is a contravariant functor that induces an equivalence between the category of varieties with dominant rational maps and the category of fields of finite transcendence degree over $\bar{k}$.*

(iii) *In particular, two varieties are birationally equivalent if and only if their function fields are isomorphic.*

PROOF. See Hartshorne [1, I.4, Theorem 4]. □

We define one more type of map. These maps play a role in algebraic geometry analogous to the role that covering maps play in topology.

Definition. Let $\phi : V \to W$ be a morphism of affine varieties, and use the map $\phi^* : \bar{k}[W] \to \bar{k}[V]$ described in (A.1.2.1(ii)) to make $\bar{k}[V]$ into a $\bar{k}[W]$-module. The morphism ϕ is called *finite* if $\bar{k}[V]$ is a finitely generated $\bar{k}[W]$-module.

A morphism $\phi : V \to W$ between varieties is *finite* if for every affine open subset $U \subset W$, the set $\phi^{-1}(U)$ is affine and the map $\phi : \phi^{-1}(U) \to U$ is finite.

Notice that a map ϕ between affine varieties is dominant if and only if ϕ^* is injective, so we say that ϕ is *finite surjective* if it is finite and ϕ^* is injective. It is true, but not obvious from the definition, that a finite map is also finite in the intuitive sense. That is, if $\phi : V \to W$ is a finite map, then it is a closed map and all fibers $\phi^{-1}(x)$ consist of a finite number of points. Further, there is an integer d and a nonempty open $U \subset \phi(V)$ such that $\#\phi^{-1}(x) = d$ for all $x \in U$. The degree d can be described algebraically as the degree of the associated field extension, and we define this quantity to be the *degree of the finite map* ϕ,

$$\deg(\phi) = \left[\bar{k}(V) : \phi^*\bar{k}(W)\right].$$

Under further hypothesis, for example that W is smooth or normal (see Section A.4 and Exercise A.1.15 for these notions), it is even true that for all $x \in \phi(V)$ we have $\#\phi^{-1}(x) \le \deg(\phi)$. However, this is not true in general; see Exercise A.1.15 for an example.

This section has been somewhat barren of examples, so we offer the following collection as a remedy. (See the exercises for more examples, especially Exercises A.1.6, A.1.7, and A.1.8). All of these examples are important tools for proving results in algebraic and arithmetic geometry.

Examples A.1.2.6. (a) (d-uple embedding) Let $M_0(x), \ldots, M_N(x)$ be the complete collection of monomials of degree d in the variables $x_0, \ldots, x_n$. Note that $N = \binom{n+d}{n} - 1$. Then the map

$$\begin{aligned} \Phi_d : \mathbb{P}^n &\longrightarrow \mathbb{P}^N, \\ x &\longmapsto \big(M_0(x), \ldots, M_N(x)\big), \end{aligned}$$

is called the *d-uple embedding of* $\mathbb{P}^n$. It is a morphism, and in fact it is actually an embedding of $\mathbb{P}^n$ into $\mathbb{P}^N$.

(b) (Segre maps) Let $m, n \ge 1$ be integers and let $N = (n+1)(m+1) - 1$. We define the *Segre map* $S_{n,m}$ by the formula

$$\begin{aligned} S_{n,m} : \mathbb{P}^n \times \mathbb{P}^m &\longrightarrow \mathbb{P}^N, \\ (x, y) &\longmapsto (x_i y_j)_{\substack{0 \le i \le n \\ 0 \le j \le m}}, \end{aligned}$$

where we have written $x = (x_0, \ldots, x_n) \in \mathbb{P}^n$ and $y = (y_0, \ldots, y_m) \in \mathbb{P}^m$. The Segre maps are again morphisms and give embeddings of the product $\mathbb{P}^n \times \mathbb{P}^m$ into $\mathbb{P}^N$. This construction explicitly displays the product of projective varieties as a projective variety.

(c) (Linear projections) Let $L_0, \ldots, L_r$ be independent linear forms in the variables $(x_0, \ldots, x_n)$, and denote by Z the linear subvariety of $\mathbb{P}^n$ defined by $L_0 = \cdots = L_r = 0$. Then we can define a rational map by the formula

$$\begin{aligned} \pi : \mathbb{P}^n &\longrightarrow \mathbb{P}^r, \\ x &\longmapsto (L_0(x), \ldots, L_r(x)). \end{aligned}$$

The domain of π clearly equals $\mathbb{P}^n \setminus Z$. We call π the *linear projection with center Z*.

(d) More generally, let $P_0, \dots, P_r \in \bar{k}[x_0, \dots, x_n]$ be homogeneous polynomials of degree d, and let Z be the algebraic subset of $\mathbb{P}^n$ defined by the equations $P_0 = \cdots = P_r = 0$. Then we can define a map

$$\begin{array}{cccc} \phi: & \mathbb{P}^n & \longrightarrow & \mathbb{P}^r \\ & x & \longmapsto & \big(P_0(x), \dots, P_r(x)\big) \end{array}$$

which is again a rational map on $\mathbb{P}^n$. Further, if the greatest common divisor of the P_i's is 1, the domain of ϕ is equal to $\mathbb{P}^n \setminus Z$. To see this, note that since homogeneous polynomials of degree d are linear combinations of monomials of degree d, it is easy to write ϕ as the composition of the d-uple embedding followed by a linear projection.

In examples (a)–(d) we wrote maps using only one chart. The next example shows that sometimes more than one chart is necessary.

(e) Let C be the curve defined in $\mathbb{P}^2$ by the equation

$$ZY^2 = X^3 + AXZ^2 + BZ^3.$$

We define a rational map $\phi : C \to \mathbb{P}^1$ by setting $\phi(X, Y, Z) = (X, Z)$. This is the restriction to C of a linear projection. The map is clearly regular except possibly at the point $(X, Y, Z) = (0, 1, 0)$. But observe that we can use the homogeneity of the coordinates and the equation for C to rewrite ϕ as

$$\begin{aligned} \phi(X, Y, Z) &= (X, Z) = (X^3, ZX^2) = (ZY^2 - AXZ^2 - BZ^3, ZX^2) \\ &= (Y^2 - AXZ - BZ^2, X^2), \end{aligned}$$

where this calculation is valid at all points on the curve with $XZ \neq 0$. This formula shows that $\phi(0, 1, 0) = (1, 0)$, and that ϕ is well-defined in a neighborhood of $(0, 1, 0)$. Thus ϕ is a morphism from C to $\mathbb{P}^1$. It is a general fact that a rational map from a smooth curve to a projective variety is always a morphism (see Theorem A.4.1.4 below).

The map ϕ is clearly finite of degree 2. In fact, $\#\phi^{-1}(P) = 2$ except for $P = (1, 0)$ and the points $P = (\alpha, 1)$ with $\alpha^3 + A\alpha + B = 0$.

(f) (Blowup of a point) Consider the projective algebraic set defined by

$$Z = \Big\{ \big((x_0, \dots, x_n), (y_0, \dots, y_{n-1})\big) \in \mathbb{P}^n \times \mathbb{P}^{n-1} \;\Big|\; x_i y_j - x_j y_i = 0 \\ \text{for all } 0 \le i \le n,\ 0 \le j \le n-1 \Big\}.$$

One can check that Z is the closure in $\mathbb{P}^n \times \mathbb{P}^{n-1}$ of the graph of the linear projection with center at $P_0 = (0, 0, \dots, 0, 1)$. Let $p : Z \to \mathbb{P}^n$

and $q : Z \to \mathbb{P}^{n-1}$ be the projections on the first and second factors, respectively, and define a map

$$\phi : \mathbb{P}^n \to \mathbb{P}^n \times \mathbb{P}^{n-1}, \qquad \phi(x_0, \dots, x_n) = ((x_0, \dots, x_n), (x_0, \dots, x_{n-1})).$$

Then it is easy to see that ϕ is a rational map from $\mathbb{P}^n$ to Z that is defined everywhere except at the point P_0. Furthermore, ϕ and p are clearly inverse to one another at every point where they are defined, so they are in fact birational maps. We observe that $p^{-1}\{P\}$ consists of the single point $\phi(P)$ except when $P = P_0$, in which case

$$p^{-1}\{P_0\} = \{P_0\} \times \mathbb{P}^{n-1}.$$

The map $p : Z \to \mathbb{P}^n$ is called the *blowup of the point P_0 on $\mathbb{P}^n$*. It has the effect of replacing the point P_0 with the projective space $\mathbb{P}^{n-1}$, while leaving all other points unchanged. In essence, the point P_0 is being replaced by the set of tangent directions through P_0.

If P_0 is a point on a projective variety V, the blowup of V at the point P_0 is defined as follows. First embed V into some $\mathbb{P}^n$ so that the image of P_0 is the point $(0, \dots, 0, 1)$. The inverse image of V by the map $p : Z \to \mathbb{P}^n$ consists of two pieces. One piece is $p^{-1}\{P_0\} = \{P_0\} \times \mathbb{P}^{n-1}$, and we denote the other piece by $\tilde{V}$. The map $p' : \tilde{V} \to V$ induced by p is called the *blowup of V at P_0*. It can be shown that this construction is independent of the chosen embedding. Clearly, the map p' is an isomorphism from $\tilde{V} \smallsetminus p'^{-1}\{P_0\}$ to $V \smallsetminus \{P_0\}$, so in particular p' is a birational morphism. For more about blowingup and some examples, see Hartshorne [1, I §4 and II §7].

(g) (Cremona transformation) The *Cremona transformation* from $\mathbb{P}^2$ to $\mathbb{P}^2$ is the rational map defined by

$$\phi(X, Y, Z) = (X^{-1}, Y^{-1}, Z^{-1}) = (YZ, XZ, XY).$$

This is readily seen to be a birational involution (i.e., $\phi \circ \phi(P) = P$ wherever it is defined), and the domain of ϕ is clearly the complement of the three points $P = (1, 0, 0)$, $Q = (0, 1, 0)$, and $R = (0, 0, 1)$. We also observe that ϕ takes the line through P, Q and sends it to the point R, and similarly for the lines through P, R and Q, R. Another, fancier, description of the Cremona transformation is to say that it first blows up the three points P, Q, R, and then it blows down the three lines $\overleftrightarrow{PQ}$, $\overleftrightarrow{PR}$, $\overleftrightarrow{QR}$.

A.1.3. Dimension

The notion of dimension is one of the most intuitive ideas in geometry, and in fact, the theorems of this section are intuitively quite clear, even if their proofs are not easy. Clearly, the dimension of a variety should be a birational invariant, and hence it is natural to define it in terms of the function field of the variety.

Definition. The *dimension* of a variety V defined over $\bar{k}$ is the transcendence degree of its function field $\bar{k}(V)$ over $\bar{k}$. The dimension of an algebraic set is the maximum of the dimensions of its irreducible components.

Not surprisingly, both $\mathbb{A}^n$ and $\mathbb{P}^n$ have dimension n. Similarly, the dimension of a hypersurface in $\mathbb{A}^n$ or $\mathbb{P}^n$ (Examples A.1.1.4 and A.1.1.6) is $n-1$. In fact, a kind of converse is true.

Proposition A.1.3.1. *A variety V of dimension $n-1$ is birational to a hypersurface in $\mathbb{A}^n$ (or $\mathbb{P}^n$).*

PROOF. This follows at once from the structure of finitely generated fields. Indeed, one can show that the field $\bar{k}(V)$ is a finite separable extension of $\bar{k}(x_1,\dots,x_{n-1})$, and so by the primitive element theorem, it is generated by a single element. See Hartshorne [1, I.4.9] for further details. □

There is another definition of dimension, which relies on the Krull dimension of a ring.

Definition. The *height of a prime ideal* $\mathfrak{p}$ in a ring A is the supremum of all n such that there exists a chain of distinct prime ideals $\mathfrak{p}_0 \subset \cdots \subset \mathfrak{p}_n = \mathfrak{p}$. The *Krull dimension* of the ring A is the supremum of the heights of its prime ideals.

The link between the Krull dimension and the geometric dimension is provided by the following theorem.

Theorem A.1.3.2. (i) *Let V be an affine algebraic set. Then*

$$\dim(V) = \text{Krulldim}(\bar{k}[V]).$$

(ii) *Let V be an affine variety and let $\mathfrak{p}$ be a prime ideal in $\bar{k}[V]$. Then*

$$\text{height}(\mathfrak{p}) + \text{Krulldim}(\bar{k}[V]/\mathfrak{p}) = \text{Krulldim}(\bar{k}[V]).$$

(iii) *Let W be a subvariety of V. Then*

$$\text{Krulldim}(\mathcal{O}_{W,V}) = \dim(V) - \dim(W).$$

PROOF. See Hartshorne [1, Theorem I.1.8.A and Exercise 3.13], Atiyah–Macdonald [1, XI] or Matsumura [1, V.14].

□

In particular, we have the following useful corollary.

Corollary A.1.3.3. *Let V be a variety, and let W be a closed algebraic subset of V. If $W \neq V$, then* $\dim W < \dim V$ *(strict inequality).*

To conform with the usual terminology, a variety of dimension one is called a *curve*, and a variety of dimension two is called a *surface*. Of course, if we are working over the field $\bar{k} = \mathbb{C}$, then a curve is also sometimes called a *Riemann surface*. Hopefully, this will not cause too much confusion.

In order to compute the dimension of a variety, we need to know how the dimension behaves for intersections of algebraic sets.

Proposition A.1.3.4. *Let V be an affine variety of dimension ℓ in $\mathbb{A}^n$, and let Z be a hypersurface in $\mathbb{A}^n$. Then either V is contained in Z, or else all of the components of $V \cap Z$ have dimension exactly $\ell - 1$. (Note that $V \cap Z$ may consist of zero components!)*

PROOF. See Shafarevich [1, I.6, Theorem 4]. □

Theorem A.1.3.5. *Let V and W be affine varieties in $\mathbb{A}^n$ of dimensions ℓ and m, respectively. Then every component of $V \cap W$ has dimension at least $\ell + m - n$.*

PROOF. The proof is by "reduction to the diagonal." First observe that $V \cap W$ is isomorphic to the intersection in $\mathbb{A}^{2n}$ of the diagonal Δ and $V \times W$. Next note that Δ is defined by n hyperplanes $x_i - y_i = 0$. So n applications of Proposition A.1.3.4 gives the theorem. □

Theorem A.1.3.6. *Let V and W be projective varieties in $\mathbb{P}^n$ of dimensions ℓ and m, respectively. Then every component of $V \cap W$ has dimension at least $\ell + m - n$. Furthermore, if $\ell + m - n \geq 0$, then $V \cap W$ is not empty.*

PROOF. Let $\bar{V}$ be the closure in $\mathbb{A}^{n+1}$ of the inverse image of V under the natural map $\mathbb{A}^{n+1} \smallsetminus \{0\} \to \mathbb{P}^{n+1}$. This is called the affine cone of V. Similarly, let $\bar{W}$ be the affine cone of W. By the previous theorem, all components of $\bar{V} \cap \bar{W}$ have dimension at least $\ell + m - n + 1$, and hence the dimension of the corresponding projective variety $V \cap W$ is at least $\ell + m - n$. Moreover, $\bar{V} \cap \bar{W}$ contains the point $0 \in \mathbb{A}^{n+1}$, so if $\ell + m - n \geq 0$, then $\bar{V} \cap \bar{W}$ will contain an affine line and $V \cap W$ will be nonempty. □

Theorem A.1.3.7. *Let $f : X \to Y$ be a surjective morphism of varieties.*

(i) $\dim(f^{-1}\{y\}) \geq \dim(X) - \dim(Y)$ *for all $y \in Y$.*

(ii) *There is a nonempty open subset $U \subset Y$ such that*

$$\dim(f^{-1}\{y\}) = \dim(X) - \dim(Y) \qquad \textit{for all } y \in U.$$

PROOF. See Shafarevich [1, I.6, Theorem 7]. The statements are true even with $f^{-1}\{y\}$ replaced by any of its irreducible components. □

Notice that the example of a blowup (Example A.1.2.6(f)) shows that the dimension of the fibers of a morphism need not be constant. As a special case of Theorem A.1.3.7, we note that if $\phi : X \to Y$ is a finite surjective morphism, then $\dim(X) = \dim(Y)$.

A.1.4. Tangent Spaces and Differentials

The purpose of this section is to define the classical differential calculus in a purely algebraic manner so that we can apply concepts like smoothness, tangent spaces, and differentials.

Let V be an affine variety defined by the equations

$$f_1(x_1, \dots, x_n) = \cdots = f_m(x_1, \dots, x_n) = 0.$$

A natural way to define the tangent space to V at the point $P = (a_1, \dots, a_n)$ is by the equations

$$\sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(P)(x_i - a_i) = 0 \qquad \text{for } 1 \le j \le m.$$

Of course, derivatives of polynomials can be defined formally over any field, without recourse to any limiting process, by repeated application of the familiar rules

$$\frac{d}{dX}(f+g) = \frac{df}{dX} + \frac{dg}{dX} \qquad \text{and} \qquad \frac{d}{dX}(aX^n) = anX^{n-1}.$$

One should perhaps also quote Leibniz's rule

$$\frac{d}{dX}(fg) = f\frac{dg}{dX} + g\frac{df}{dX}.$$

To see that the definition of the tangent space is intrinsic, that is, independent of the particular defining equations for V, we give another definition, which is valid for arbitrary varieties.

Definition. Let P be a point on a variety V. The *tangent space to V at P* is the $\bar{k}$-vector space

$$T_P(V) = \operatorname{Hom}_{\bar{k}}(\mathcal{M}_{P,V}/\mathcal{M}_{P,V}^2, \bar{k}).$$

In other words, the tangent space is defined to be the dual of the vector space $\mathcal{M}_{P,V}/\mathcal{M}_{P,V}^2$. We naturally call $\mathcal{M}_{P,V}/\mathcal{M}_{P,V}^2$ the *cotangent space to V at P*. It is easy to see that the tangent and cotangent spaces are $\bar{k}$-vector spaces, since $\mathcal{O}_{P,V}/\mathcal{M}_{P,V} \cong \bar{k}$. It is also not hard to check that this definition agrees with the naive definition; see, for example, Shafarevich [1, II.1, Theorem 1] or Mumford [6, III.4]. We also note that the tangent and cotangent spaces are defined at every point of V, not only at the "nonsingular points." In fact, we will use the tangent space in the definition of nonsingularity.

Theorem A.1.4.1. *Let V be a variety. Then $\dim(T_P(V)) \geq \dim(V)$ for all $P \in V$. Furthermore, there exists a nonempty open subset $U \subset V$ such that $\dim(T_P(V)) = \dim(V)$ for all $P \in U$.*

PROOF. See Hartshorne [1, I.5, Proposition 2A and Theorem 3] or Shafarevich [1, II.1, Theorem 3]. □

Definition. A point P on a variety V is *singular* if $\dim(T_P(V)) > \dim(V)$, and it is *nonsingular* (or *smooth*) if $\dim(T_P(V)) = \dim(V)$. The variety V is called *nonsingular* or *smooth* if all of its points are nonsingular.

We see from Theorem A.1.4.1 that a variety always has an open subset of smooth points. The following criterion is frequently used to compute the singular points of a variety.

Lemma A.1.4.2. (Jacobian criterion) *Let V be an affine variety defined by the equations*

$$f_1(x_1, \dots, x_n) = \cdots = f_m(x_1, \dots, x_n) = 0,$$

and let $P = (a_1, \dots, a_n)$ be a point on V. Then P is a smooth point if and only if

$$\operatorname{Rank}\left(\frac{\partial f_j}{\partial x_i}(P)\right)_{\substack{1\le j\le m\\ 1\le i\le n}} = n - \dim(V).$$

PROOF. See Hartshorne [1, I.5] or Mumford [6, III.4, Corollary 1]. □

Consider a rational map $f : V \to W$ that is regular at P, and let $Q = f(P)$. We have seen that f induces a homomorphism of local rings, $f^* : \mathcal{O}_{Q,W} \to \mathcal{O}_{P,V}$, and hence it induces a $\bar{k}$-linear map

$$f^* : \mathcal{M}_{Q,W}/\mathcal{M}_{Q,W}^2 \longrightarrow \mathcal{M}_{P,V}/\mathcal{M}_{P,V}^2,$$

which we again denote by f^*.

Definition. The *tangent map* $df(P) : T_P(V) \to T_Q(W)$ is the transpose of the map $f^* : \mathcal{M}_{Q,W}/\mathcal{M}_{Q,W}^2 \to \mathcal{M}_{P,V}/\mathcal{M}_{P,V}^2$.

Theorem A.1.4.3. *Let V be a variety and let $P \in V$ be a smooth point. Then the local ring $\mathcal{O}_{P,V}$ is a regular local ring.*

PROOF. This is clear from the definitions and Theorem A.1.3.2(i), since $\mathcal{O}_{P,V}$ is regular if the dimension of $\mathcal{M}_{P,V}/\mathcal{M}_{P,V}^2$ is equal to the Krull dimension of $\mathcal{O}_{P,V}$. □

Theorem A.1.4.4. *Let ϕ be a rational map from a smooth variety V to a projective variety. Then*

$$\operatorname{codim}_V\bigl(V \smallsetminus \operatorname{dom}(\phi)\bigr) \geq 2.$$

In other words, a rational map on a smooth variety is defined except possibly on a set of codimension at least 2.

PROOF. See Shafarevich [1,III.3, Theorem 3] or Silverman [2, IV.6.2.1]. □

We now discuss the theory of differential forms on a variety X. Probably the right language to use is that of sheaves, but for the moment we will take a more concrete approach. The starting point is the differential of a function $f \in k(X)^*$. For any point $x \in \operatorname{dom}(f)$ we have a tangent map $df(x) : T_x(X) \to T_{f(x)}(\mathbb{A}^1) = k$, so $df(x)$ is a linear form on $T_x(X)$. We note that the classical rules $d(f+g) = df + dg$ and $d(fg) = f\,dg + g\,df$ are valid. Thus we may view df as a map that associates to each point $x \in \operatorname{dom}(f)$ a linear form on $T_x(X)$ (i.e., a cotangent vector). We call such a map an *abstract differential form*, but of course we need to impose some sort of continuity condition as x varies. So we take all of the abstract differential forms that can be built up out of the df's.

Definition. A *regular differential* 1*-form* on a variety X is an abstract differential form ω such that for all $x \in X$ there is a neighborhood U of x and regular functions $f_i, g_i \in \mathcal{O}(U)$ such that $\omega = \sum f_i\,dg_i$ on U. We denote the set of regular 1-forms on X by $\Omega^1[X]$. It is clearly a k-vector space, and in fact, it is an $\mathcal{O}(X)$-module.

Examples A.1.4.5.
(1) The space of regular differential 1-forms on affine space $\mathbb{A}^n$ is

$$\Omega^1[\mathbb{A}^n] = \bigoplus_{i=1}^{n} k[t_1, \ldots, t_n]\,dt_i,$$

where $t_1, \ldots, t_n$ are affine coordinates for $\mathbb{A}^n$. Indeed, $k[\mathbb{A}^n] = k[t_1, \ldots, t_n]$, and the differentials of polynomials clearly belong to and generate this space.
(2) Let ω be a regular differential 1-form on $\mathbb{P}^n$. Then on any $\mathbb{A}^n \subset \mathbb{P}^n$, it must have the shape $\omega = \sum_{i=1}^{n} P_i(t)dt_i$. However, if any of the P_i's are nonzero, ω will have poles along the hyperplane at infinity, so it will not be regular. Therefore, $\Omega^1[\mathbb{P}^n] = 0$.

So we see that global 1-forms behave quite differently from local ones. We next want to write $\Omega^1[U]$ as a direct sum in a manner analogous to the description for $\Omega^1[\mathbb{A}^n]$ in A.1.4.5.

Definition. Let x be a nonsingular point on a variety X of dimension n. Functions $t_1, \ldots, t_n \in \mathcal{O}_x$ are called *local parameters* at x if the t_i's are in $\mathcal{M}_x$ and if they give a basis of $\mathcal{M}_x/\mathcal{M}_x^2$. The functions $t_1, \ldots, t_n$ give *local coordinates* on X if $t'_i := t_i - t_i(x)$ give local parameters at all x in X.

Recall that $\mathcal{M}_x/\mathcal{M}_x^2$ is dual to the tangent space, so local parameters exist only at nonsingular points. It is easy to see that $t_1, \ldots, t_n$ are local parameters if and only if $\bigcap_i \ker(dt_i(x)) = \{0\}$ in $T_x(X)$, and thus that local parameters give local coordinates on a neighborhood of x. From Nakayama's lemma one may deduce the following result.

Proposition A.1.4.6. *Let x be a nonsingular point on X. Then there exist local parameters $t_1, \ldots, t_n$ at x and a neighborhood U of x such that $\Omega^1[U] = \oplus_{i=1}^n \mathcal{O}(U) dt_i$.*

PROOF. See Shafarevich [1, III.4 Theorem 1]. □

So far, we have considered only the tangent vector spaces $T_x(X)$ and their duals, but we may also construct their exterior powers $\bigwedge^r T_x(X)^*$ and copy the classical definition of differential forms of higher order. (Recall that $\bigwedge^r V$ is the space of r-linear skew-symmetric forms on the vector space V.)

Definition. An abstract r-form ω on a variety X assigns to each $x \in X$ a linear map $\omega(x) : \bigwedge^r T_x(X) \to k$. A *regular r-form* ω on X is an abstract r-form such that for all $x \in X$ there is a neighborhood U containing x and functions $f_i, g_{i_1,\ldots,i_r} \in \mathcal{O}(U)$ such that

$$\omega = \sum g_{i_1,\ldots,i_r} df_{i_1} \wedge \cdots \wedge df_{i_r}.$$

We will let $\Omega^r[U]$ denote the space of regular r-forms on U. It is clearly an $\mathcal{O}(U)$-module. The analogue of Proposition A.1.4.6 is true. If $t_1, \ldots, t_n$ are local coordinates on U, then

$$\Omega^r[U] = \bigoplus_{i_1 < \cdots < i_r} \mathcal{O}(U) dt_{i_1} \wedge \cdots \wedge dt_{i_r}.$$

For some examples of computation with differentials see Exercises A.1.10, A.2.7, and A.4.2.(f) and Theorem A.4.2.6 and the remarks following it.

It is convenient to define a *rational differential form* to be a form that is regular on an open subset, where we identity two differential forms if they coincide on some open subset. The space of such forms is denoted by $\Omega^r(X)$ and is clearly a vector space of dimension $\binom{n}{r}$ over $k(X)$, where $n = \dim(X)$.

Definition. Let $\phi : X \to Y$ be a morphism of smooth varieties. Then there is a map $\phi^* : \Omega^r[Y] \to \Omega^r[X]$ defined by the formula

$$\phi^*\Bigl(\sum g_{i_1,\dots,i_r} df_{i_1} \wedge \cdots \wedge df_{i_r}\Bigr) = \sum (g_{i_1,\dots,i_r} \circ \phi) d(f_{i_1} \circ \phi) \wedge \cdots \wedge d(f_{i_r} \circ \phi).$$

As usual, the spaces of differential forms are functorial and contravariant, that is, $(\phi\circ\psi)^*(\omega) = \psi^*(\phi^*(\omega))$. One of the features that makes spaces of differential forms important is they can be used to define invariants of a variety X.

Lemma A.1.4.7. *Let $\phi : X \dashrightarrow Y$ be a dominant rational map between smooth projective varieties. Then $\phi^*(\Omega^r[Y]) \subset \Omega^r[X]$. In other words, the poles of ϕ do not create any poles for the differential form $\phi^*(\omega)$. Hence if ϕ is a birational map, then $\Omega^r[X]$ and $\Omega^r[Y]$ are isomorphic.*

PROOF. See Shafarevich [1, III.5, Theorem 2]. □

For example, we will see that if X is a (smooth) projective variety of dimension n, then $g(X) := \dim \Omega^n[X]$ is finite. This follows from Corollary A.3.2.7 below, and in fact, every $\Omega^r[X]$ is finite-dimensional. The quantity $g(X)$ is called the *geometric genus* of X. We will study it in detail for curves in Section A.4.

We close this section with some material on algebraic groups.

Definition. An *algebraic group* defined over k is a variety G defined over k, a point $e \in G(k)$, and morphisms $m : G \times G \to G$ and $i : G \to G$ satisfying the axioms of a group law:

(i) $m(e,x) = m(x,e) = x$.
(ii) $m\bigl(i(x),x\bigr) = m\bigl(x,i(x)\bigr) = e$.
(iii) $m\bigl(m(x,y),z\bigr) = m\bigl(x,m(y,z)\bigr)$.

Remark. Sometimes this definition of algebraic group is relaxed to include reducible sets with a group law. Then the irreducible components are disjoint and form a finite group, which we denote by $\Phi(G)$. The connected component of G containing e, denoted by G^0, is then an algebraic group in the above sense; we call it the *identity component of* G.

Using the definitions, one sees that for any $g \in G$, the right and left translation maps

$$R_g : \begin{array}{ccc} G & \longrightarrow & G, \\ h & \longmapsto & m(g,h), \end{array} \qquad \text{and} \qquad L_g : \begin{array}{ccc} G & \longrightarrow & G, \\ h & \longmapsto & m(h,g), \end{array}$$

are isomorphisms. From this remark we deduce that algebraic groups are smooth varieties. Indeed, if there were a singular point, then using the translation maps and their tangent maps, we would deduce that all points are singular, contradicting Theorem A.1.4.1. It is also easy to see that the tangent map at the origin associated to the group operation m,

$$dm(e,e) : T_{(e,e)}(G \times G) = T_e(G) \times T_e(G) \longrightarrow T_e(G),$$

is just the addition of vectors in $T_e(G)$.

Examples A.1.4.8. (a) The *additive group* $\mathbb{G}_a$ is the variety $\mathbb{A}^1$ with the group law being addition:

$$m : \mathbb{G}_a \times \mathbb{G}_a \longrightarrow \mathbb{G}_a, \qquad m(x, y) = x + y.$$

(b) The *multiplicative group* $\mathbb{G}_m$ is the variety $\mathbb{A}^1 \setminus \{0\}$ with the group law being multiplication:

$$m : \mathbb{G}_m \times \mathbb{G}_m \longrightarrow \mathbb{G}_m, \qquad m(x, y) = xy.$$

(c) The *general linear group* $\mathrm{GL}(n)$ is the group of $n \times n$ invertible matrices with the group law being matrix multiplication. Note that although $\mathrm{GL}(n)$ is naturally defined as the quasi-projective variety

$$\mathrm{GL}(n) = \left\{(x_{ij}) \in \mathbb{A}^{n^2} \mid \det(x_{ij}) \neq 0\right\},$$

it is actually an affine variety, since we can also define it as

$$\mathrm{GL}(n) = \left\{(x_{ij}, t) \in \mathbb{A}^{n^2} \times \mathbb{A}^1 \mid t \det(x_{ij}) = 1\right\}.$$

It is known that every affine algebraic group is a subgroup of $\mathrm{GL}(n)$ for some n. It is harder to give examples of algebraic groups that are not affine. We will see in Section A.4 that smooth plane cubics are algebraic groups, called elliptic curves. In fact, elliptic curves and their higher-dimensional analogues, abelian varieties, will be one of our main objects of study in this book.

Definition. An *abelian variety* is a projective variety that is also an algebraic group.

Although it is far from obvious from looking at the definition, it can be shown that the group law on an abelian variety is necessarily commutative. (See Lemma A.7.1.3). For perspective, we quote the following structure theorem.

Theorem A.1.4.9. (Chevalley) *Let G be an algebraic group defined over k. There exists a maximal connected affine subgroup H of G. This subgroup H is defined over k and is a normal subgroup of G. The quotient of G by H has a natural structure as an abelian variety.*

PROOF. See Rosenlicht [1, Theorem 16]. □

EXERCISES

A.1.1. (a) Let V be a variety that is both affine and projective. Prove that V consists of a single point.
(b) Let V be a projective variety, let W be an affine variety, and let $\phi : V \to W$ be a regular map. Prove that ϕ is constant.

A.1.2. Let X and Y be projective varieties defined over a field k.

(a) If $X(k) \neq \emptyset$ and there is a k-morphism $f : X \to Y$, prove that $Y(k) \neq \emptyset$.

(b) If X and Y are k-birationally equivalent, prove that $X(k)$ is dense in X (for the Zariski topology) if and only if $Y(k)$ is dense in Y.

(c) Prove that if X has a smooth k-rational point and if there exists a rational map from X to Y defined over k, then $Y(k) \neq \emptyset$. (*Hint.* Take well-chosen hyperplane sections and use induction on the dimension to reduce to the case that X is a curve.) Deduce that if X and Y are smooth and k-birationally equivalent, then $X(k) \neq \emptyset$ if and only if $Y(k) \neq \emptyset$. (For the necessity of the smoothness assumption, see the next exercise.)

A.1.3. (a) Show that the Cremona map $(x, y, z) \mapsto (x^{-1}, y^{-1}, z^{-1})$ on $\mathbb{P}^2$ gives a birational isomorphism between the two curves C and C' defined by $x^2 + y^2 = az^2$ and $y^2z^2 + x^2z^2 = ax^2y^2$.

(b) Let $a \in k^*$ and assume that $\operatorname{char}(k) \neq 2$. Verify that C is smooth and that C' has three singular points, namely $(0 : 0 : 1)$, $(0 : 1 : 0)$, and $(1 : 0 : 0)$.

(c) Show that if $k = \mathbb{Q}$, the set $C(\mathbb{Q})$ is empty for some values of a (e.g., for $a = 3$). Conclude that the property of having a k-rational point is not a birational property of (singular) varieties, even in dimension 1.

A.1.4. (a) Show that over an algebraically closed field, a smooth conic is isomorphic to the projective line. (*Hint.* See Section A.4.3 below.)

(b) Show that the rings $k[X, Y]$ and $k[X_0, X_1, X_2]/(X_0^2 - X_1X_2)$ are not isomorphic, and conclude that the homogeneous coordinate ring of a variety $V \subset \mathbb{P}^n$ is not an invariant of V. In other words, the homogeneous coordinate ring depends on the projective embedding. (*Hint.* Show that the second ring is not a unique factorization domain.)

A.1.5. Let $f : X \to Y$ be a morphism of affine varieties.

(a) Prove that f^* is injective if and only if f is dominant (i.e., $f(X)$ is dense in Y).

(b) Prove that f^* is surjective if and only if f is a closed embedding (i.e., $f(X)$ is a closed subvariety of Y and $f : X \to f(X)$ is an isomorphism).

(c) Show that $V = \mathbb{A}^2 \smallsetminus \{(0, 0)\}$ is not an affine variety. (*Hint.* Show that the injection of V into $\mathbb{A}^2$ induces an isomorphism between $k[\mathbb{A}^2]$ and $\mathcal{O}(V)$ and use Theorem A.1.2.1 to derive a contradiction.)

(d) Show that the only regular functions on $X = \mathbb{P}^2 \smallsetminus \{(1, 0, 0)\}$ are the constants. Deduce that X is neither affine nor projective.

A.1.6. Let R and Q be homogeneous polynomials of degree 2, and let V be the smooth cubic surface in $\mathbb{P}^3$ defined by

$$x_0Q(x_2, x_3) - x_1R(x_2, x_3) = 0.$$

(a) Show that $\phi(x_0, \ldots, x_3) = (x_0, x_1)$ defines a morphism from V to $\mathbb{P}^1$.

(b) Show the same for $\psi(x_0, \ldots, x_3) = (x_2, x_3)$.

(c) Prove that the map $\phi \times \psi : V \to \mathbb{P}^1 \times \mathbb{P}^1$ is a birational morphism.

(d) Prove that such a birational morphism exists for any smooth cubic surface containing two skew lines. (In fact, two such lines always exist, although they may be defined only over an extension of k.)

A.1.7. (Frobenius map) In this exercise we work with the finite fields $\mathbb{F}_{p^r}$ containing p^r elements. Let V be a quasi-projective variety defined over $\mathbb{F}_{p^r}$, and let F be the map $F(x_0, \dots, x_n) := (x_0^p, \dots, x_n^p)$.
(a) Show that F maps V onto another quasi-projective variety $V^{(p)}$ defined over $\mathbb{F}_{p^r}$.
(b) If V is projective (respectively affine), prove that $V^{(p)}$ is also projective (respectively affine).
(c) Prove that $F : V \to V^{(p)}$ is a bijection on points, but that it is not an isomorphism of varieties.
(d) Suppose that V is defined over $\mathbb{F}_p$. Prove that $V^{(p)} = V$ and that

$$V(\mathbb{F}_{p^r}) = \{x \in V(\bar{\mathbb{F}}_p) \mid F^r(x) = x\}.$$

Here $F^r = F \circ F \circ \cdots \circ F$ is the r-fold iterate of F.

A.1.8. (a) Show that the map $f : (x_0, x_1) \mapsto (x_0^3, x_0x_1^2, x_1^3)$ is a morphism from $\mathbb{P}^1$ to $\mathbb{P}^2$. Show that the image of f is a projective curve C and find its equation. Is the map $f : \mathbb{P}^1 \to C$ a bijection on points? Is it birational? Is it an isomorphism?
(b) Answer the same questions for the map $g : \mathbb{P}^1 \to \mathbb{P}^2$ defined by the formula $g : (x_0, x_1) \mapsto (x_0^3, x_0(x_1^2 - x_0^2), x_1(x_1^2 - x_0^2))$.

A.1.9. (Resolution of the singularities of a map). Let $f : X \dashrightarrow Y$ be a rational map between varieties. Show that there exists a variety $\tilde{X}$ with a morphism $\tilde{f} : \tilde{X} \to Y$ and a birational morphism $p : \tilde{X} \to X$ such that $\tilde{f} = f \circ p$. Furthermore, show that we may impose the condition that $p : p^{-1}(\operatorname{dom}(f)) \to \operatorname{dom}(f)$ is an isomorphism. (*Hint.* Choose $\tilde{X}$ equal to the closure of the graph of f in $X \times Y$, and take p and $\tilde{f}$ to be the projections onto X and Y.)

A.1.10. For each of the following varieties X, describe the space of regular r-forms $\Omega^r[X]$ for each $0 \le r \le \dim X$.
(a) $X = \mathbb{A}^n$.
(b) $X = \mathbb{P}^n$.
(c) $X \subset \mathbb{P}^2$ is the smooth projective cubic curve $x^3 + y^3 + z^3 = 0$. (*Hint.* Show that $\Omega^1[X]$ is a vector space of dimension 1.)
(d) Let η be a third root of unity and define $\phi(x, y, z) = (x, \eta y, z)$. Check that ϕ is an automorphism of the curve in (c) and compute $\phi^*(\omega)$ for any 1-form ω.

A.1.11. (Grassmannian varieties) Consider $V = \mathbb{A}^{n+1}$ as a vector space of dimension $n + 1$, and let $\mathbb{P}V = \mathbb{P}^n$ be its associated projective space. Let $\operatorname{Gras}(k, \mathbb{P}V) = \operatorname{Gras}(k, n)$ be the set of all linear subspaces of dimension k in $\mathbb{P}V = \mathbb{P}^n$, or equivalently, the set of vector subspaces of dimension $k+1$ in $\mathbb{A}^{n+1}$. We want to give $\operatorname{Gras}(k, n)$ the structure of a projective algebraic variety.
(a) Let W be a subspace of V of dimension $k + 1$ and select a basis $w_0, \dots, w_k$. Then the (multi)vector $w_0 \wedge \cdots \wedge w_k$ is a nonzero element of $\bigwedge^{k+1} V$ and thus defines a point in $\mathbb{P}(\bigwedge^{k+1} V)$. Show that the map thus defined from $\operatorname{Gras}(k, \mathbb{P}V)$ to $\mathbb{P}(\bigwedge^{k+1} V)$ is well-defined (i.e., independent of choice of basis) and injective. Denote by X the image of this map in $\mathbb{P}(\bigwedge^{k+1} V)$.

(b) Let $\omega \in \bigwedge^{k+1} V$, and show that the map $\delta(\omega) : v \mapsto \omega \wedge v$ from V to $\bigwedge^{k+2} V$ has rank less than $n-k$ if and only if $\omega \in X$. Conclude that X is a projective algebraic set and thus that we may endow $\mathrm{Gras}(k,n)$ with the structure of an algebraic variety. (*Hint.* The entries of a matrix describing $\delta(\omega)$ are linear forms, and X will be defined by the vanishing of some minors of this matrix.)

(c) Recall that there is a natural (but unique only up to scalars) isomorphism between $\bigwedge^{k+1} V$ and $\bigwedge^{n-k} V^*$, where * denotes the dual space. We denote one such isomorphism by $\omega \mapsto \omega^*$. For $\omega \in \bigwedge^{k+1} V$ we get a map $\delta'(\omega) : v^* \mapsto v^* \wedge \omega^*$ from V^* to $\bigwedge^{n-k+1} V^*$. Show that ω is in X if and only if $\delta'(\omega)$ has rank at most $k+1$, and that the subspace W corresponding to ω is the orthogonal complement of $\ker \delta'(\omega)$. That is, the transpose maps ${}^t\delta(\omega) : \bigwedge^{k+2} V^* \to V^*$ and ${}^t\delta'(\omega) : \bigwedge^{n-k} V \to V$ have orthogonal images.

(d) Deduce from part (c) that $\omega \in \mathbb{P}(\bigwedge^{k+1} V)$ is in X if and only if

$$\langle {}^t\delta(\omega)(x^*), {}^t\delta'(\omega)(x) \rangle = 0 \text{ for all } x^* \in \textstyle\bigwedge^{k+2} V^*,\ x \in \bigwedge^{n-k} V.$$

Deduce that X is cut out by quadratic forms.

(e) Notice that $\mathrm{Gras}(0,n) = \mathbb{P}^n$ and $\mathrm{Gras}(n-1,n) = \mathbb{P}V^* \cong \mathbb{P}^n$. Show that $\mathrm{Gras}(1,3)$, the variety of lines in $\mathbb{P}^3$, is a quadric in $\mathbb{P}^5$.

(f) Show that there is a transitive action of $\mathrm{GL}(n+1)$ on $\mathrm{Gras}(k,n)$ defined by the map $(f,W) \mapsto f(W)$. Deduce that $\mathrm{Gras}(k,n)$ is an irreducible smooth variety of dimension $(k+1)(n-k)$. (*Hint.* To find the dimension, compute the dimension of the fibers of the map $f \mapsto f(W)$ for a fixed W.)

Remark. The embedding described in this exercise is classically called the *Plücker embedding*, the coordinates on $\mathbb{P}(\bigwedge^{k+1} V)$ are called *Plücker coordinates*, and the quadratic forms of part (d) are called *Plücker relations*. It can be shown that the Plücker relations generate the ideal defining $\mathrm{Gras}(k,n)$ inside $\mathbb{P}(\bigwedge^{k+1} V)$.

A.1.12. Let x be a smooth point on X, and let $u_1, \dots, u_n$ be local parameters. Recall that this means that $u_i \in \mathcal{M}_x \subset \mathcal{O}_{x,X}$ and $u_1, \dots, u_n$ generate $\mathcal{M}_x/\mathcal{M}_x^2$ as a k-vector space.

(a) Show that for all regular functions $f \in \mathcal{O}_{x,X}$ there is a unique polynomial S_m of degree at most m such that $f - S_m(u_1, \dots, u_n) \in \mathcal{M}_x^{m+1}$.

(b) Show that $S(f) := \lim S_m$ exists in $k[\![X_1, \dots, X_n]\!]$ and that the resulting map $S : \mathcal{O}_{x,X} \to k[\![X_1, \dots, X_n]\!]$ is an injective ring homomorphism. (*Hint.* Use the fact that $\bigcap_m \mathcal{M}_x^m = \{0\}$.)

(c) The ring $k[\![X_1, \dots, X_n]\!]$ has a natural topology given by the powers of the ideal generated by $X_1, \dots, X_n$, and with this topology, $k[\![X_1, \dots, X_n]\!]$ is complete. One can similarly endow $\mathcal{O}_{x,X}$ with the topology defined by the powers of $\mathcal{M}_x$. Show that S is continuous with dense image, and conclude that if $\widehat{\mathcal{O}}_{x,X}$ denotes the completion of $\mathcal{O}_{x,X}$, then S extends to an isomorphism $S : \widehat{\mathcal{O}}_{x,X} \to k[\![X_1, \dots, X_n]\!]$.

A.1.13. (a) Let k be a perfect field, and let X be an affine (or projective) variety defined over $\bar{k}$. Prove that the following three conditions are equivalent:

(i) X is the set of common zeros of polynomials with coefficients in k.

(ii) The ideal defining X is generated by polynomials with coefficients in k.

(iii) X is globally invariant under the action of $\mathrm{Gal}(\bar{k}/k)$. That is, if $x \in X(\bar{k})$ and $\sigma \in \mathrm{Gal}(\bar{k}/k)$, then $\sigma(x) \in X(\bar{k})$.

(b) Give an example to show that (a) need not be true over a nonperfect field such as $k = \mathbb{F}_p(T)$.

A.1.14. A variety is said to be *complete* or *proper over* k if every projection $X \times Y \to Y$ is closed (i.e., the image of every closed subset is closed).

(a) Prove that projective varieties are complete.

(b) Show that a regular function f on a complete variety must be constant (*Hint.* Consider the closed set $\{(x,t) \in X \times \mathbb{A}^1 \mid f(x)t = 1\}$ and its projection to $\mathbb{A}^1$.)

A.1.15. A variety X is called *normal* at a point x if the local ring $\mathcal{O}_{x,X}$ is integrally closed. The variety X is *normal* if it is normal at every point.

(a) Prove that a smooth variety is normal.

(b) If X is affine, show that X is normal if and only if $k[X]$ is integrally closed. Assuming $\mathrm{char}(k) \neq 2$, use this to prove the normality of the affine cone X in $\mathbb{A}^3$ given by the equation $x^2 + y^2 - z^2 = 0$. (Notice that X is not smooth.)

(c) Let X be an affine variety with coordinate ring R, and let R' be the integral closure of R, so R' is a finitely generated k-algebra. (Prove this yourself or see Zariski–Samuel [1, Chapter V, Theorem 9].) Thus R' corresponds to a normal affine variety X', and there is a morphism $\nu : X' \to X$. Show that ν is a finite surjective birational morphism with the following universal property: For any normal variety Z and any dominant morphism $\phi : Z \to X$, there is a unique morphism $\phi' : Z \to X'$ such that $\phi = \nu \circ \phi'$. Intuitively, X' is the "smallest" normal variety that maps onto X. (*Hint.* The integral closure of $\phi^* k[X]$ sits inside $k[Z]$ and is isomorphic to $k[X']$.) The variety X' is called the *normalization of* X.

(d) Show that a curve is normal if and only it is smooth. Hence normalization provides a method to resolve the singularities of a curve.

A.1.16. Let $P \in \mathbb{P}^n$. Prove that P has homogeneous coordinates $(x_0, \ldots, x_n)$ with all $x_i \in k$ if and only if $\sigma(P) = P$ for all $\sigma \in G_k$. (*Hint.* You will need to use Hilbert's theorem 90, $H^1(G_k, \bar{k}^*) = 0$.)

A.1.17. Let V be a closed subvariety of dimension r in $\mathbb{P}^n$. Let $\check{\mathbb{P}}^n$ denote the projective space dual to $\mathbb{P}^n$ and identify points in $\check{\mathbb{P}}^n$ with hyperplanes in $\mathbb{P}^n$. We set

$$Z_V = \{(x; H_0, \ldots, H_r) \in V \times \left(\check{\mathbb{P}}^n\right)^{r+1} \mid x \in H_0 \cap \ldots \cap H_r\}.$$

Compute the dimension of Z_V and conclude that the set

$$Y_V = \{(H_0, \ldots, H_r) \in \left(\check{\mathbb{P}}^n\right)^{r+1} \mid V \cap H_0 \cap \ldots \cap H_r \neq \emptyset\}$$

is a hypersurface in $\left(\check{\mathbb{P}}^n\right)^{r+1}$. The multihomogeneous form F_V that defines Y_V is unique up to a scalar; it is called the *Chow form* of V. If V' is another subvariety of the same dimension, show that $F_{V'}$ is a scalar multiple of F_V if and only if $V = V'$. (Hint: Use the dimension theorems).

A.2. Divisors

A polynomial in one variable is determined up to a scalar by its roots, counted with multiplicities. A polynomial in several variables is determined, again up to a scalar, by the hypersurfaces counted with multiplicities on which it vanishes. Further, these hypersurfaces with their multiplicities correspond exactly to the decomposition of the polynomial into irreducible factors. The theory of divisors is a device that generalizes this idea to arbitrary varieties, where unique factorization no longer holds. We will look at two ways of defining divisors. The first, due to Weil, is as a sum of subvarieties of codimension one. The second, due to Cartier, is as objects that are locally defined by one equation. Weil's definition is more concrete and works well on normal varieties, but Cartier's definition is frequently easier to work with and yields a better theory on nonnormal varieties and more general schemes.

Throughout most of this section we work over an algebraically closed field k. It will not be until we get to Proposition A.2.2.10 that we will explain why everything we have done carries over to arbitrary perfect fields. This proposition is easy, but it will be fundamental for our further work.

A.2.1. Weil Divisors

As indicated above, a Weil divisor is a sum of subvarieties of codimension one.

Definition A.2.1.1. Let X be an algebraic variety. The *group of Weil divisors on* X is the free abelian group generated by the closed subvarieties of codimension one on X. It is denoted by $\operatorname{Div}(X)$.

In other words, a divisor is a finite formal sum of the form $D = \sum n_Y Y$, where the n_Y's are integers and the Y's are codimension-one subvarieties of X. For example, if X is a curve, then the Y's are points; if X is a surface, then the Y's are (irreducible) curves; and so on.

The *support of the divisor* $D = \sum n_Y Y$ is the union of all those Y's for which the multiplicity n_Y is nonzero. It is denoted by $\operatorname{supp}(D)$. The divisor is said to be *effective* or *positive* if every $n_Y \geq 0$.

We recall that if Y is an irreducible divisor on X, then $\mathcal{O}_{Y,X}$ is the local ring of functions regular in a neighborhood of some point of Y. In particular, if the variety X is nonsingular, or more generally if it is nonsingular along Y, then $\mathcal{O}_{Y,X}$ is a discrete valuation ring. We write

$\text{ord}_Y : \mathcal{O}_{Y,X} \setminus \{0\} \to \mathbb{Z}$ for the normalized valuation on $\mathcal{O}_{Y,X}$, and we extend ord_Y to the fraction field $k(X)^*$ in the usual way. We refer the reader to Fulton [1] for the definition of ord_Y in the general case. The following lemma (see also Exercise A.2.3) summarizes its main properties.

Lemma A.2.1.2. *The order function* $\text{ord}_Y : k(X)^* \to \mathbb{Z}$ *described above has the following properties:*

(i) $\text{ord}_Y(fg) = \text{ord}_Y(f) + \text{ord}_Y(g)$ *for all* $f, g \in k(X)^*$.

(ii) *Fix* $f \in k(X)^*$. *There are only finitely many* Y*'s with* $\text{ord}_Y(f) \neq 0$.

(iii) *Let* $f \in k(X)^*$. *Then* $\text{ord}_Y(f) \geq 0$ *if and only if* $f \in \mathcal{O}_{Y,X}$. *Similarly,* $\text{ord}_Y(f) = 0$ *if and only if* $f \in \mathcal{O}_{Y,X}^*$.

(iv) *Assume further that* X *is projective, and let* $f \in k(X)^*$. *Then the following are equivalent:*

(a) $\text{ord}_Y(f) \geq 0$ *for all* Y.
(b) $\text{ord}_Y(f) = 0$ *for all* Y.
(c) $f \in k^*$.

The properties of ord_Y described in Lemma A.2.1.2 allow us to define the divisor of a function.

Definition. Let X be a variety, and let $f \in k(X)^*$ be a rational function on X. The *divisor of* f is the divisor

$$\text{div}(f) = \sum_Y \text{ord}_Y(f) Y \in \text{Div}(X).$$

A divisor is said to be *principal* if it is the divisor of a function. Two divisors D and D' are said to be *linearly equivalent*, denoted by $D \sim D'$, if their difference is a principal divisor. For brevity, we also sometimes write (f) for the divisor of f.

The *divisor of zeros of* f, denoted by $(f)_0$, and the *divisor of poles of* f, denoted by $(f)_\infty$, are defined by

$$(f)_0 = \sum_{\text{ord}_Y(f)>0} \text{ord}_Y(f) Y \qquad \text{and} \qquad (f)_\infty = \sum_{\text{ord}_Y(f)<0} -\text{ord}_Y(f) Y.$$

Thus the divisor of a function is the difference of its zeros and its poles (counted with the appropriate multiplicities).

Let us briefly comment on the origin of the term linear equivalence. Suppose that $D \sim D'$, say $D' = D + \text{div}(f)$. For each point $(a, b) \in \mathbb{P}^1$, define a divisor $D_{(a,b)} := D + \text{div}(a + bf)$. The divisors $D_{(a,b)}$ are parametrized by the points of the line $\mathbb{P}^1$, and clearly $D_{(1,0)} = D$ and $D_{(0,1)} = D'$. So there is a family of divisors, parametrized by the points of a line, that deforms D to D'.

Definition. The *divisor class group of* X is the group of divisor classes modulo linear equivalence. It is denoted by $\text{Cl}(X)$. The linear equivalence class of a divisor D will be denoted by $\text{Cl}(D)$.

As an important example, we can compute $\mathrm{Cl}(\mathbb{P}^n)$.

Proposition A.2.1.3. *Let* $\deg(Z)$ *denote the degree of an irreducible hypersurface* $Z \subset \mathbb{P}^n$*, and extend the function* deg *by linearity to the group of divisors* $\mathrm{Div}(\mathbb{P}^n)$*. Then a divisor* $D \in \mathrm{Div}(\mathbb{P}^n)$ *is principal if and only if it has degree* 0*, and the induced map*

$$\deg : \mathrm{Cl}(\mathbb{P}^n) \to \mathbb{Z}$$

is an isomorphism.

PROOF. A hyperplane $H \subset \mathbb{P}^n$ has degree $\deg(H) = 1$, so the degree map is surjective. We must show that its kernel consists of exactly the principal divisors.

Let $f \in k(\mathbb{P}^n)^*$. We can write $f = P/Q$ with P, Q homogeneous polynomials of degree d. We factor P and Q into irreducible factors as $P = P_1^{m_1} \cdots P_r^{m_r}$ and $Q = Q_1^{n_1} \cdots Q_s^{n_s}$, and we set $Y_i = Z(P_i)$ and $Z_j = Z(Q_j)$. Thus the Y_i's and Z_j's are irreducible hypersurfaces in $\mathbb{P}^n$, and it is clear from the definitions that $\mathrm{div}(f) = \sum_{i=1}^r m_i Y_i - \sum_{j=1}^s n_j Z_j$. Further,

$$\deg\bigl(\mathrm{div}(f)\bigr) = \sum_{i=1}^{r} m_i \deg(Y_i) - \sum_{j=1}^{s} n_j \deg(Z_j) = \deg P - \deg Q = 0,$$

which shows that principal divisors have degree zero.

Conversely, suppose that D is a divisor with $\deg(D) = 0$. Then we can write D as $D = \sum_{i=1}^r m_i Y_i - \sum_{j=1}^s n_j Z_j$, where $\sum m_i \deg(Y_i) = \sum n_j \deg(Z_j)$. Let P_i be an irreducible homogeneous polynomial defining Y_i, and similarly let Q_j define Z_j, and set $P = P_1^{m_1} \cdots P_r^{m_r}$ and $Q = Q_1^{n_1} \cdots Q_s^{n_s}$. Then

$$\begin{aligned} \deg P &= \sum_{i=1}^{r} m_i \deg(P_i) = \sum_{i=1}^{r} m_i \deg(Y_i) \\ &= \sum n_j \deg(Z_j) = \sum n_j \deg(Q_j) = \deg Q. \end{aligned}$$

Therefore, $f = P/Q$ is in $k(\mathbb{P}^n)$, and clearly $\mathrm{div}(f) = D$, which proves that every divisor of degree zero is principal. □

Remark A.2.1.4. By the same method, one can show that

$$\mathrm{Cl}\left(\mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_r}\right) = \mathbb{Z}^r.$$

See Exercise A.2.1.

Remark A.2.1.5. Consider the natural map $k(X)^* \to \mathrm{Div}(X)$ that takes a function to its divisor. Lemma A.2.1.2(iv) says that if X is projective, then the kernel of this map consists of only the constant functions. So on a projective variety, the divisor class group fits into (and can be defined by) the exact sequence

$$0 \longrightarrow k^* \longrightarrow k(X)^* \xrightarrow{\mathrm{div}} \mathrm{Div}(X) \longrightarrow \mathrm{Cl}(X) \longrightarrow 0.$$

This sequence is the analogue of the exact sequence relating the unit group R_K^* and the ideal class group Cl_K of a number field K,

$$0 \longrightarrow R_K^* \longrightarrow K^* \longrightarrow \{\text{Fractional ideals}\} \longrightarrow \mathrm{Cl}_K \longrightarrow 0.$$

If X is affine and $A = k[X]$ is integrally closed, the analogy can be pushed even further, since then $\mathrm{Cl}(X) = 0$ if and only if A is a UFD, i.e., a unique factorization domain. (See Exercise A.2.8).

A.2.2. Cartier Divisors

A subvariety of codimension one on a normal variety is defined locally as the zeros and poles of a single function. The idea of a Cartier divisor is to take this local property as the definition, subject to the condition that the functions fit together properly.

Definition. A *Cartier divisor* on a variety X is an (equivalence class of) collections of pairs $(U_i, f_i)_{i \in I}$ satisfying the following conditions:

(i) The U_i's are open sets that cover X.
(ii) The f_i's are nonzero rational functions $f_i \in k(U_i)^* = k(X)^*$.
(iii) $f_i f_j^{-1} \in \mathcal{O}(U_i \cap U_j)^*$ (i.e., $f_i f_j^{-1}$ has no poles or zeros on $U_i \cap U_j$).

Two collections $\{(U_i, f_i) \mid i \in I\}$ and $\{(V_j, g_j) \mid j \in J\}$ are considered to be equivalent (define the same divisor) if $f_i g_j^{-1} \in \mathcal{O}(U_i \cap V_j)^*$ for all $i \in I$ and $j \in J$.

The sum of two Cartier divisors is

$$\{(U_i, f_i) \mid i \in I\} + \{(V_j, g_j) \mid j \in J\} = \{(U_i \cap V_j, f_i g_j) \mid (i, j) \in I \times J\}.$$

With this operation, the Cartier divisors form a group that we denote by $\mathrm{CaDiv}(X)$. The *support* of a Cartier divisor is the set of zeros and poles of the f_i's. A Cartier divisor is said to be *effective* or *positive* if it can be defined by a collection $\{(U_i, f_i) \mid i \in I\}$ with every $f_i \in \mathcal{O}(U_i)$. (That is, f_i has no poles on U_i.)

Associated to a function $f \in k(X)^*$ is its Cartier divisor, denoted by

$$\operatorname{div}(f) = \{(X, f)\}.$$

Such a divisor is called a *principal Cartier divisor.* Two divisors are said to be *linearly equivalent* if their difference is a principal divisor. The group of Cartier divisor classes modulo linear equivalence is called the *Picard group of* X and is denoted by $\operatorname{Pic}(X)$. (In many texts the Picard group is defined as the group of line bundles or invertible sheaves on X. We will generally be working with varieties, for which these two groups coincide.)

Remark. The reader with more knowledge of sheaf cohomology will recognize that a Cartier divisor is nothing more than a global section of the quotient sheaf $\mathcal{K}_X^*/\mathcal{O}_X^*$. The principal divisors are those that come from global sections of $\mathcal{K}^*$, and one can show that the group of Cartier divisor classes $\operatorname{Pic}(X)$ is isomorphic to the cohomology group $H^1(X, \mathcal{O}_X^*)$. For more details, see Hartshorne [1, II §6] and the introduction to Section A.3.3.

We now compare the two types of divisors. Let Y be an irreducible subvariety of codimension 1 in X, and let D be a Cartier divisor defined by $\{(U_i, f_i) \mid i \in I\}$. We define the order of D along Y, denoted by $\operatorname{ord}_Y(D)$, as follows. Select one of the open subsets U_i such that $U_i \cap Y \neq \emptyset$ and set $\operatorname{ord}_Y(D) = \operatorname{ord}_Y(f_i)$. It is easily seen that $\operatorname{ord}_Y(D)$ is independent of the choice of (U_i, f_i), so we obtain a map from Cartier divisors to Weil divisors by sending D to $\sum \operatorname{ord}_Y(D)Y$. Clearly, this map sends effective Cartier divisors to effective Weil divisors and principal Cartier divisors to principal Weil divisors, and hence it induces a map from $\operatorname{Pic}(X)$ to $\operatorname{Cl}(X)$. In general, this map is neither surjective nor injective. For example, see Fulton [2, Examples 2.1.2 and 2.1.3] or Hartshorne [1, II.6.11.3]. However, there are a number of important cases for which it is a bijection, including the one described in the following theorem.

Theorem A.2.2.1. *Let X be a smooth variety. Then the natural maps*

$$\operatorname{CaDiv}(X) \longrightarrow \operatorname{Div}(X) \qquad \textit{and} \qquad \operatorname{Pic}(X) \longrightarrow \operatorname{Cl}(X)$$

are isomorphisms.

PROOF. See Hartshorne [1, II.6.11]. In fact, it suffices to assume that the local rings of X are unique factorization domains. □

In the sequel we will consider only Cartier divisors when the variety in question might be singular, and we will freely identify Weil and Cartier divisors when we work with smooth varieties.

Example A.2.2.2. *The Divisor Cut Out by a Hypersurface.* Let $X \hookrightarrow \mathbb{P}^n$ be a projective variety, let I_X be its homogeneous ideal, and let F be a homogeneous polynomial of degree d not in I_X. Note that this means

that F does not vanish identically on X. We want to associate to F a divisor $(F)_X$ that corresponds to the intersection of X and the zero locus F. To do this, we cover X by the affine open subsets $U_i = X \smallsetminus \{x_i = 0\}$, and then $(F)_X$ is defined by the collection

$$(F)_X = \{(U_i, F/x_i^d) \mid 0 \leq i \leq n\}.$$

Notice that F/x_i^d is a well-defined rational function on U_i, and that the ratio $(F/x_i^d)(F/x_j^d)^{-1} = (x_j/x_i)^d$ has no zeros or poles on $U_i \cap U_j$, so these functions patch together to give a Cartier divisor. Further, F/x_i^d clearly has no poles on U_i, so the divisor $(F)_X$ is effective. Finally, we observe that if G is any other homogeneous polynomial of degree d not in I_X, then F/G is a rational function on X and

$$(F)_X - (G)_X = \operatorname{div}(F/G),$$

so $(F)_X$ and $(G)_X$ are linearly equivalent. In this way we obtain a natural injection $\mathbb{Z} \hookrightarrow \operatorname{Pic}(X)$ associated to the embedding $X \hookrightarrow \mathbb{P}^n$.

Example A.2.2.3. *The Canonical Divisor on a Smooth Variety.* Let X be a smooth variety of dimension n, and let ω be a nonzero differential n-form on X. We can construct a divisor associated to ω as follows. On any affine open subset U of X with local coordinates $x_1, \ldots, x_n$ we can write $\omega = f_U \, dx_1 \wedge \cdots \wedge x_n$ for some rational function $f_U \in k(X)$. We then define the divisor of ω by the collection

$$\operatorname{div}(\omega) = \{(U, f_U)\}.$$

Taking different affine coordinates will give an equivalent collection, and the Jacobian transformation formula for differential forms shows that the pairs (U, f_U) patch together to produce a well-defined divisor on X.

Any other nonzero differential n-form ω' on X has the form $\omega' = f\omega$ for some rational function $f \in k(X)^*$. It follows that

$$\operatorname{div}(\omega') = \operatorname{div}(\omega) + \operatorname{div}(f),$$

so the divisor class associated to an n-form is independent of the chosen form. This divisor class is called the *canonical class of* X. It is an extremely important invariant of the variety X. By abuse of language, any divisor in the canonical class is called a *canonical divisor* and is denoted by K_X. We also observe that though we cannot speak of the value of a differential form at some point, it makes sense to say that it vanishes at some point.

Example A.2.2.4. *The Canonical Divisor on Projective Space.* Let h be the divisor class of a hyperplane in projective space $\mathbb{P}^n$. Then the canonical divisor on $\mathbb{P}^n$ is given by $K_{\mathbb{P}^n} \sim -(n+1)h$. See Exercise A.2.4.

In general, objects constructed in mathematics are useful only if they have some functoriality properties. For this reason, we will attach to each morphism (and even each rational map) between varieties a map between their Picard groups. In this way the association $X \mapsto \operatorname{Pic}(X)$ will become a contravariant functor.

Definition. Let $g : X \to Y$ be a morphism of varieties, let $D \in \operatorname{CaDiv}(Y)$ be a Cartier divisor defined by $\{(U_i, f_i) \mid i \in I\}$, and assume that $g(X)$ is not contained in the support of D. Then the Cartier divisor $g^*(D) \in \operatorname{CaDiv}(X)$ is the divisor defined by

$$g^*(D) = \{(g^{-1}(U_i), f_i \circ g) \mid i \in I\}.$$

It is immediate from the definition that $g^*(D+E) = g^*(D) + g^*(E)$ whenever they are defined, and that if $g : X \to Y$ and $f : Y \to Z$ are two morphisms of varieties, then $(f \circ g)^* = g^* \circ f^*$. It is also clear that

$$g^*(\operatorname{div}(f)) = \operatorname{div}(f \circ g),$$

provided that the rational function $f \in k(Y)$ gives a well-defined rational function on $g(X)$. However, unless f is a dominant map (i.e., unless the image is dense), then $f^*(D)$ will not be well-defined for all D. The next lemma allows us to move D in its linear equivalence class, and thereby to define f^* on all of $\operatorname{Pic}(X)$.

Lemma A.2.2.5. *Let $f : X \to Y$ be a morphism of varieties.*

(i) *Let $D, D' \in \operatorname{CaDiv}(Y)$ be linearly equivalent divisors. If $f(X)$ is not contained in $\operatorname{supp}(D) \cup \operatorname{supp}(D')$, then $f^*(D) \sim f^*(D')$.*

(ii) (Moving lemma) *For every Cartier divisor $D \in \operatorname{CaDiv}(Y)$ there exists another Cartier divisor $D' \in \operatorname{CaDiv}(Y)$ satisfying*

$$D \sim D' \qquad \text{and} \qquad f(X) \not\subset \operatorname{supp}(D').$$

PROOF. We are given that $D' = D + \operatorname{div}(g)$. If D is defined by $\{(U_i, f_i)\}$, then D' is defined by $\{(U_i, f_i g)\}$. Hence $f^*(D)$ and $f^*(D')$ are defined respectively by $\{(f^{-1}U_i, f_i \circ f)\}$ and $\{(f^{-1}U_i, (f_i g) \circ f)\}$, which shows that $f^*D' = f^*D + \operatorname{div}(g \circ f)$. This proves (i).

To prove the moving lemma, we again let D be defined by $\{(U_i, f_i)\}$. For any fixed index $j \in I$ we define a divisor D_j by $\{(U_i, f_i f_j^{-1})\}$. Then $D_j = D - \operatorname{div}(f_j)$, so $D_j \sim D$, and clearly $U_j \cap \operatorname{supp}(D_j) = \emptyset$. □

As an immediate consequence of Lemma A.2.2.5, we obtain the following result.

Proposition A.2.2.6. *Let $f : X \to Y$ be a morphism of varieties. The map $f^* : \operatorname{CaDiv}(Y) \to \operatorname{CaDiv}(X)$, which is well-defined only for D such that $f(X) \not\subset \operatorname{supp}(D)$, induces a (well-defined) homomorphism $f^* : \operatorname{Pic}(Y) \to \operatorname{Pic}(X)$.*

Example A.2.2.7. (a) Let $\Phi_d : \mathbb{P}^n \to \mathbb{P}^N$ be the d-uple embedding (A.1.2.6(a)). Using the isomorphism $\operatorname{Pic}(\mathbb{P}^N) = \mathbb{Z}$ described in Proposition A.2.1.3, the map $\Phi_d^* : \operatorname{Pic}(\mathbb{P}^N) \to \operatorname{Pic}(\mathbb{P}^n)$ is simply the multiplication-by-d map $\mathbb{Z} \to \mathbb{Z}$, $z \mapsto dz$.
(b) Let $S : \mathbb{P}^m \times \mathbb{P}^n \to \mathbb{P}^N$ be the Segre map (A.1.2.6(b)). Using the isomorphisms $\operatorname{Pic}(\mathbb{P}^m \times \mathbb{P}^n) = \mathbb{Z}^2$ and $\operatorname{Pic}(\mathbb{P}^N) = \mathbb{Z}$ from Proposition A.2.1.3 and the remark following it, the map $S^* : \operatorname{Pic}(\mathbb{P}^N) \to \operatorname{Pic}(\mathbb{P}^m \times \mathbb{P}^n)$ is the diagonal map $\mathbb{Z} \to \mathbb{Z}^2$, $z \mapsto (z, z)$.
(c) Let $i : X \hookrightarrow \mathbb{P}^n$ be an embedding of a projective variety X into projective space. Then the divisor $(F)_X$ on X cut out by a hypersurface $Z(F)$ (see A.2.2.2) is equal to $i^*\big(Z(F)\big)$.
(d) Let $f \in \bar{k}(X)$ be a nonconstant function on a smooth curve X. Then f defines a morphism $\bar{f} : X \to \mathbb{P}^1$, and clearly

$$\bar{f}^*\big((0) - (\infty)\big) = \operatorname{div}(f).$$

More generally, if X is any smooth variety and $f \in \bar{k}(X)$ any nonconstant function, then f extends to a rational map $\bar{f} : X \to \mathbb{P}^1$ that is well-defined except possibly on a set of codimension at least 2. Hence we can define $\bar{f}^*\big((0) - (\infty)\big)$, and again we find that it is equal to $\operatorname{div}(f)$.

As an illustration of the general theory, we will compute the effect of f^* on the canonical class of a variety when f is a finite map. To do this, we need to measure the ramification of f along a divisor.

Definition. Let $f : X \to Y$ be a finite map of smooth projective varieties, let Z be an irreducible divisor on X, and let $Z' = f(Z)$ be the image of Z under f. Note that the dimension theorem (A.1.3.7) tells us that Z' is an irreducible divisor on Y. Let s_Z be a generator of the maximal ideal of $\mathcal{O}_{Z,X}$, and similarly let $s_{Z'}$ be a generator of the maximal ideal of $\mathcal{O}_{Z',Y}$. (That is, s_Z and $s_{Z'}$ are local equations for Z and Z'.) The *ramification index of f along Z* is defined to be the integer

$$e_Z = e_Z(f) = \operatorname{ord}_Z(s_{Z'} \circ f),$$

where we recall that $\operatorname{ord}_Z : \mathcal{O}_{Z,X} \to \mathbb{Z}$ is the valuation on $\mathcal{O}_{Z,X}$. Equivalently, $s_{Z'} \circ f = u s_Z^{e_Z}$ for some function $u \in \mathcal{O}_{Z,X}^*$. The map f is said to be *ramified along Z* if $e_Z(f) \geq 2$.

We now investigate how the pullback of the canonical class on Y compares with the canonical class on X.

Proposition A.2.2.8. (Hurwitz formula) *Let $f : X \to Y$ be a finite map between smooth projective varieties.*

(i) *The map f is ramified only along a finite number of irreducible divisors.*

(ii) *If we assume further either that the characteristic of k is 0 or that the characteristic of k does not divide any of the ramification indices, then we have the formula*

$$K_X \sim f^*(K_Y) + \sum_Z \bigl(e_Z(f) - 1\bigr) Z.$$

PROOF. The proof of the first assertion will follow from the proof of the second one. We prove the formula (ii). Let $n = \dim(X) = \dim(Y)$, and choose an n-form ω on Y. We will compare $\operatorname{div}\bigl(f^*(\omega)\bigr)$ and $f^*\bigl(\operatorname{div}(\omega)\bigr)$. Let Z be an irreducible divisor on X, and let $e = e_Z(f)$ be its ramification index. Fix local coordinates $y_1, \dots, y_n$ on Y so that $t = y_1$ is a local equation for $Z' = f(Z)$ (this may require shrinking a little bit the open set on which we work). We select local coordinates $x_1, \dots, x_n$ on X as follows: $s = x_1$ is a local equation for Z, and $x_i = y_i \circ f$ for $i = 2, \dots, n$. We write

$$\omega = \phi(x)\, dy_1 \wedge \cdots \wedge dy_n \quad \text{and} \quad f^*(\omega) = (\phi \circ f)\,(dy_1 \circ f) \wedge \cdots \wedge (dy_n \circ f).$$

We know that $t \circ f = us^e$ for some function u that does not have a zero or a pole along Z, so we get $dy_1 \circ f = es^{e-1}ds + s^e du$. Notice that the hypothesis on the characteristic implies $e \neq 0$ in k, so $e \in k^*$. Hence we obtain

$$f^*(\omega) = \phi \circ f s^{e-1} u'\, dx_1 \wedge \cdots \wedge dx_n \quad \text{with } u' \in \mathcal{O}^*_{Z,X}.$$

We conclude that $\operatorname{ord}_Z\bigl(f^*(\omega)\bigr) = \operatorname{ord}_Z\bigl(f^*\bigl(\operatorname{div}(\omega)\bigr)\bigr) + (e_Z - 1)$. Now, both $\operatorname{ord}_Z\bigl(f^*(\omega)\bigr)$ and $\operatorname{ord}_Z\bigl(f^*\bigl(\operatorname{div}(\omega)\bigr)\bigr)$ are zero except for a finite number of divisors Z; this proves (i), and Hurwitz's formula follows by summing over all divisors Z. □

To each divisor D we associate the vector space of rational functions whose poles are no worse than D. The precise definition is as follows.

Definition. Let D be a divisor on a variety X. The vector space $L(D)$ is defined to be the set of rational functions

$$L(D) = \bigl\{ f \in k(X)^* \mid D + \operatorname{div}(f) \geq 0 \bigr\} \cup \{0\}.$$

The dimension of $L(D)$ as a k-vector space is denoted by $\ell(D)$.

To check that $L(D)$ is a vector space, we use the fact that ord_Y is a valuation. Thus if $f, g \in L(D)$, then for any irreducible divisor Y we have

$$\operatorname{ord}_Y(f+g) \geq \min\{\operatorname{ord}_Y(f), \operatorname{ord}_Y(g)\} \geq -\operatorname{ord}_Y(D).$$

Summing over Y shows that $f+g \in L(D)$. It will be shown in the next section that the dimension $\ell(D)$ is finite when X is projective. We also note the following elementary properties, whose proof we will leave to the reader (see Exercise A.2.5).

Lemma A.2.2.9. *Let X be a variety and let $D, D' \in \operatorname{Div}(X)$.*
(i) *$k \subset L(D)$ if and only if $D \geq 0$.*
(ii) *If $D \leq D'$, then $L(D) \subset L(D')$.*
(iii) *If $D' = D + \operatorname{div}(g)$, then the map $f \mapsto gf$ gives an isomorphism of k-vector spaces $L(D') \to L(D)$. In particular, the dimension $\ell(D)$ depends only on the class of D in $\operatorname{Pic}(X)$.*

We close this section on divisors by explaining what happens when the field k is not assumed to be algebraically closed. So for the rest of this section we drop the assumption that k is algebraically closed and assume for simplicity only that k is perfect. (If one needs to work with nonperfect fields such as $\mathbb{F}_q(T)$, one should use the separable closure instead of the algebraic closure in the following discussion.) We first need to explain what it means for a divisor to be defined over k. We do this by using the action of the Galois group $G_k := \operatorname{Gal}(\bar{k}/k)$.

Definition. Let X be a variety defined over k. A divisor D is said to be *defined over k* if it is invariant under the action of the Galois group G_k.

For example, a hypersurface $X \subset \mathbb{P}^n$ that is defined over k is a divisor defined over k. Similarly, the principal divisor $\operatorname{div}(f)$ of a rational function $f \in k(X)$ is defined over k.

If the divisor D is defined over k, we consider the k-vector space $L_k(D)$ defined by

$$L_k(D) = \{f \in k(X) \mid D + \operatorname{div}(f) \geq 0\}.$$

The next proposition clarifies the connection between $L_k(D)$ and $L(D) = L_{\bar{k}}(D)$ and justifies the assertion that for questions concerning divisors, it generally suffices to work over $\bar{k}$.

Proposition A.2.2.10. *Let k be a perfect field, let $\bar{k}$ be an algebraic closure of k, let X be a variety defined over k, and let D be a divisor on X defined over k.*
(i) *The $\bar{k}$-vector space $L_{\bar{k}}(D) = L(D)$ has a basis of elements in $k(X)$. In other words, there is a natural identification $L_k(D) \otimes_k \bar{k} = L(D)$, and in particular, $\dim L_k(D) = \dim L(D) = \ell(D)$.*
(ii) *Assume further that X is projective. If there exists a rational function $f \in \bar{k}(X)$ with $D = \operatorname{div}(f)$, then there exists a rational function $f' \in$*

$k(X)$ with $D = \operatorname{div}(f')$. In other words, the natural map from $\operatorname{Pic}(X)_k$ *to* $\operatorname{Pic}(X)$ *is injective.*

PROOF. (i) It suffices to prove that every element of $L(D)$ can be expressed as a $\bar{k}$-linear combination of elements of $k(X)$. Let $f \in L(D)$, and let K/k be a finite Galois extension such that $f \in L_K(D)$. Further let $\alpha_1, \ldots, \alpha_n$ be a basis for K over k, and let $\operatorname{Gal}(K/k) = \{\sigma_1, \ldots, \sigma_n\}$. It is a basic fact of Galois theory that $\det(\sigma_i(\alpha_j)) \neq 0$; see, for example, Lang [2, Chapter VIII, Corollary 5.4]. Define rational functions g_i by

$$g_i = \sum_{j=1}^{n} \sigma_j(\alpha_i f), \qquad 1 \leq i \leq n.$$

It is easy to check that the g_i's are $\operatorname{Gal}(K/k)$-invariant, and thus are in $k(X)$. And the invertibility of the matrix $(\sigma_i(\alpha_j))$ shows that f (and in fact, all of the $\sigma_j f$'s) are in the $\bar{k}$-span of $g_1, \ldots, g_n$.
(ii) By assumption, $D = \operatorname{div}(f)$ for some $f \in \bar{k}(X)$. We fix a finite Galois extension K/k such that $f \in K(X)$. The fact that $\sigma(D) = D$ for all $\sigma \in \operatorname{Gal}(K/k)$ means that $\operatorname{div}(\sigma(f)/f) = 0$. It follows from Lemma A.2.1.4(iv) that $\sigma(f)/f$ is constant, say $\sigma(f)/f = a(\sigma) \in K^*$. One easily verifies that the map $\sigma \to a(\sigma)$ is a one-cocycle from $\operatorname{Gal}(K/k)$ to K^*. Hilbert's theorem 90 (Serre [1, Chapter 10, Proposition 2] or Exercise A.2.6) tells us that it is a coboundary, so there is a $b \in K^*$ such that $a(\sigma) = b \cdot \sigma(b)^{-1}$ for all $\sigma \in \operatorname{Gal}(K/k)$. It follows that $\sigma(bf) = bf$ for all $\sigma \in \operatorname{Gal}(K/k)$, so $bf \in k(X)$. Since we clearly have $\operatorname{div}(bf) = \operatorname{div}(f) = D$, this completes the proof. □

A.2.3. Intersection Numbers

A classical part of algebraic geometry called enumerative geometry is dedicated to counting the number of points (or curves, etc.) satisfying certain properties. We will introduce one basic tool used in studying this kind of problem. By the general theorems on dimensions (reviewed in Section A.1.3) we expect that a collection of n hypersurfaces on a variety of dimension n will intersect in a finite set of points. We would like to count these points, including some sort of multiplicity to account for tangencies and self-intersections. We begin by defining the intersection multiplicity of n irreducible divisors $D_1, \ldots, D_n$ at a point $x \in X$ under the assumption that $\bigcap_i D_i$ consists of discrete points.

Definition. Let X be a variety of dimension n, and let $D_1, \ldots, D_n \in \operatorname{Div}(X)$ be irreducible divisors with the property that $\dim(\bigcap_i D_i) = 0$.

Choose local equations $f_1, \ldots, f_n$ for $D_1, \ldots, D_n$ in a neighborhood of a point $x \in X$. The *(local) intersection index of* $D_1, \ldots, D_n$ *at* x is

$$(D_1, \ldots, D_n)_x = \dim_k \left(\mathcal{O}_{x,X}/(f_1, \ldots, f_n) \right).$$

One can check that this dimension is always finite and does not depend on the selected local equations. Further, it is positive if and only if $x \in \bigcap_i D_i$. We define the *intersection index (or number) of* $D_1, \ldots, D_n$ to be

$$(D_1, \ldots, D_n) = \sum_{x \in X} (D_1, \ldots, D_n)_x.$$

If the D_i's intersect transversally, that is, if each $(D_1, \ldots, D_n)_x$ is either 0 or 1, then $(D_1, \ldots, D_n)$ actually counts the number of points in $D_1 \cap \cdots \cap D_n$. We can easily extend this definition by linearity so as to define the intersection number of any n divisors, as long as their intersection consists only of points. To go further, we need the following invariance property.

Lemma A.2.3.1. *Let* X *be a normal projective variety and* $D_1, \ldots, D_n \in \operatorname{Div}(X)$.
(i) *There exist divisors* $D'_1, \ldots, D'_n \in \operatorname{Div}(X)$ *with the property that*

$$D_i \sim D'_i \text{ for all } 1 \le i \le n \qquad \text{and} \qquad \left(\bigcap_{i=1}^{n} \operatorname{supp}(D'_i) \right) = 0.$$

(ii) *Let* $D'_1, \ldots, D'_n$ *be as in* (i), *and suppose that* $D_1, \ldots, D_n$ *also satisfy* $\dim\left(\bigcap_i \operatorname{supp}(D_i)\right) = 0$. *Then the intersection numbers are equal,*

$$(D_1, \ldots, D_n) = (D'_1, \ldots, D'_n).$$

PROOF. The first part can easily be proven in the same way as the moving lemma A.2.2.5(ii). For the second part, see Shafarevich [1, IV.1, Theorem 2]. □

Lemma A.2.3.1 enables us to define the intersection number of any n-uple of divisors.

Definition. Let X be a normal projective variety of dimension n. For any divisors $D_1, \ldots, D_n$ on X, choose divisors $D'_1, \ldots, D'_n \in \operatorname{Div}(X)$ as in (A.2.3.1(i)). Thus $D_i \sim D'_i$ for all $1 \le i \le n$ and $\dim\left(\bigcap_i \operatorname{supp}(D'_i)\right) = 0$. We define the *intersection index (or number) of* $D_1, \ldots, D_n$ to be

$$(D_1, \ldots, D_n) = (D'_1, \ldots, D'_n).$$

Note that Lemma A.2.3.1(ii) assures us that this number is independent of the choice of $D'_1, \ldots, D'_n$.

The following important theorem explains how intersection indices transform under finite morphisms.

Theorem A.2.3.2. *Let X and Y be normal projective varieties of dimension n, and let $f : X \to Y$ be a finite morphism. Let $D_1, \ldots, D_n \in \operatorname{Div}(Y)$. Then*

$$(f^*D_1, \ldots, f^*D_n)_X = \deg(f) \cdot (D_1, \ldots, D_n)_Y.$$

PROOF. See Mumford [2, II.6] or Exercise A.2.10. □

Finally, we define the degree of a subvariety with respect to a divisor.

Definition. Let X be a projective variety, let $i : Z \hookrightarrow X$ be a subvariety of dimension r, and let $D \in \operatorname{Div}(X)$. The *degree of Z with respect to D* is defined to be

$$\deg_D(Z) = \big(\underbrace{i^*(D), \ldots, i^*(D)}_{r \text{ times}}\big)_Z.$$

When $X = \mathbb{P}^n$ and $D = H$ is a hyperplane, the degree of Z with respect to H is called the *projective degree* of Z. In this case, the degree is just the number of points in the intersection of Z with a general linear subvariety of codimension r. For example, the degree of a hypersurface with respect to a hyperplane is the degree of the polynomial that defines the hypersurface. Similarly, the degree of a point with respect to a hyperplane is one.

Remark. (i) The intersection numbers actually satisfy a much stronger invariance property than that of Lemma A.2.3.1, although we will not need to use this fact. They are invariant by algebraic deformation; that is, the intersection number does not change if each divisor is changed to an algebraically equivalent divisor. We recall that two divisors D_1, D_2 on X are *algebraically equivalent* if there exists a connected algebraic set T, two points $t_1, t_2 \in T$, and a divisor $\mathcal{D}$ on $X \times T$ such that $D_i = \mathcal{D}|_{X \times \{t_i\}}$ for $i = 1, 2$. See, for example, Hartshorne [1, V, Exercise 1.7] or Griffiths–Harris [1, pages 461–2].
(ii) Although the degree is defined with respect to any divisor, it is a useful concept only when the divisor is ample (a notion introduced in the next section). In this case the degree has a simple geometric interpretation; see Exercise A.3.6.

EXERCISES

A.2.1. (a) Show that an (irreducible) hypersurface Y in $\mathbb{P} := \mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_r}$ is defined by an (irreducible) multihomogeneous polynomial of multidegree $(d_1, \ldots, d_r)$.
(b) Use linearity to define a map $\deg : \operatorname{Div}(\mathbb{P}) \to \mathbb{Z}^r$ and show that it induces an isomorphism between $\operatorname{Pic}(\mathbb{P})$ and $\mathbb{Z}^r$.

A.2.2. (a) Show that an automorphism f of $\mathbb{P}^n$ must transform a hyperplane into a hyperplane.
(b) Let H be a hyperplane in $\mathbb{P}^n$. Using the action of f on $L(H)$, conclude that all automorphisms of $\mathbb{P}^n$ are linear. In other words, prove that $\mathrm{Aut}(\mathbb{P}^n) = \mathrm{PGL}(n+1)$.

A.2.3. Prove in detail Lemma A.2.1.2. For part (ii), first reduce to an affine open subset U on which f is regular, and then show that $\mathrm{ord}_Y(f) > 0$ if and only if Y is contained in the closed subset of U defined by the ideal $f \cdot k[U] \subset k[U]$. For part (iv) use the fact that the only regular functions on a projective variety are constants.

A.2.4. Compute the canonical class of $\mathbb{P}^n$. (*Hint.* Use the differential $\omega := dx_1 \wedge \cdots \wedge dx_n$, where $x_i = X_i/X_0$, and show that $\mathrm{div}(\omega) = -(n+1)H_0$ for a certain hyperplane H_0.)

A.2.5. (a) Verify Lemma A.2.2.9, and prove that if D, D' are two divisors on a variety X, then there is a well-defined map

$$\mu : L(D) \otimes L(D') \longrightarrow L(D+D'), \qquad \mu : (f, f') \longmapsto ff'.$$

Show that in general this map is neither surjective nor injective. (See Exercise A.3.8 for more on the map μ.)
(b) Let X, Y be smooth varieties with canonical divisor classes K_X, K_Y. Prove that $K_{X\times Y} = p_1^*(K_X) + p_2^*(K_Y)$, where p_1, p_2 are the projections from $X \times Y$ to X and Y. Use this to compute the canonical class on $\mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_r}$.

A.2.6. Prove Hilbert's theorem 90: Let K/k be a finite Galois extension and let $a : \mathrm{Gal}(K/k) \to K^*$ be a map such that $a(\sigma\tau) = \sigma(a(\tau))a(\sigma)$. Prove that there exists a $b \in K^*$ such that $a(\sigma) = b/\sigma(b)$. (*Hint.* Consider $b = \sum_{\sigma\in\mathrm{Gal}(K/k)} a(\sigma)\sigma(u)$, and show that $u \in K$ can be chosen so that $b \neq 0$.)

A.2.7. (a) Let X be a smooth hypersurface in $\mathbb{P}^n$ defined by a homogeneous polynomial F of degree d. Compute the canonical class K_X of X. (*Hint.* Use the forms

$$\omega_i := (-1)^i \frac{dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_n}{(\partial F/\partial x_i)(1, x_1, \ldots, x_n)}$$

to compute the canonical class.)
(b) Generalize (a) to the case that X is a codimension r smooth complete intersection of r hypersurfaces defined by polynomials F_i of degree d_i. (*Hint.* For I a subset of cardinality r of $[1, n]$, define

$$\Delta_I(x) = \det\left((\partial F_j/\partial x_i)(1, x_1, \ldots, x_n)\right)_{\substack{i\in I\\ 1\le j\le r}}$$

and $d\hat{x}_I = dx_1 \wedge \cdots \wedge d\hat{x}_i \wedge \cdots \wedge dx_n$ with each dx_i deleted when $i \in I$; then use the forms $\omega_I = d\hat{x}_I/\Delta_I(x)$ to show that K_X is $d_1 + \cdots + d_r - 1 - n$ times a hyperplane section.)

A.2.8. Let X be a smooth affine variety, and let $A = \mathcal{O}(X)$ be its coordinate ring. Show that $\mathrm{Cl}(X) = 0$ if and only if A is a UFD.

A.2.9. (a) Let $p : X \times \mathbb{A}^n \to X$ be projection onto the first factor. Prove that the map $p^* : \mathrm{Cl}(X) \to \mathrm{Cl}(X \times \mathbb{A}^n)$ is an isomorphism, and similarly with Pic instead of Cl.

(b) Can you describe $\mathrm{Cl}(X \times \mathbb{P}^n)$ or $\mathrm{Pic}(X \times \mathbb{P}^n)$?

A.2.10. Let $f : X \to Y$ be a finite morphism of degree d between projective varieties of dimension n. The purpose of this exercise is to furnish a proof of the formula

$$(f^*D_1, \dots, f^*D_n)_X = d(D_1, \dots, D_n)_Y$$

stated in Theorem A.2.3.2.

(a) Reduce to the case where the divisors D_i are effective and $\bigcap_{i=1}^n D_i$ is finite and avoids the ramification locus. (Use linearity, invariance of intersection numbers, and the moving lemma.)

(b) Suppose that the differential df_x is injective for some $x \in X$, assume $y = f(x)$, and let $f_1, \dots, f_n$ be local equations for $D_1, \dots, D_n$ at y. Prove that $f_1 \circ f, \dots, f_n \circ f$ are local equations for $f^*D_1, \dots, f^*D_n$ at x, and that

$$\dim_k(\mathcal{O}_x/(f_1 \circ f, \dots, f_n \circ f)) = \dim_k(\mathcal{O}_y/(f_1, \dots, f_n)).$$

(c) Conclude the proof.

A.2.11. Recall the construction of the Grassmannian variety $\mathrm{Gras}(k,n)$ of linear subspaces of dimension k in $\mathbb{P}V = \mathbb{P}^n$, and identify $\mathrm{Gras}(k,n)$ with the image of its Plücker embedding into $\mathbb{P}(\bigwedge^{k+1} V)$ (see Exercise A.1.11). Let Z be a linear subspace of dimension $n-k$. Then Z corresponds to a (multi)vector $\omega \in \bigwedge^{n-k} V \cong \bigwedge^{k+1} V^*$; hence it gives a linear form on $\mathbb{P}(\bigwedge^{k+1} V)$. Let U' be the affine subset defined by the nonvanishing of this form, and let $U = U' \cap \mathrm{Gras}(k,n)$.

(a) Show that U is isomorphic to the affine space $\mathbb{A}^{(k+1)(n-k)}$ and that $\mathrm{Gras}(k,n)$ is covered by such open subsets. (*Hint.* Show that U can be identified with $\mathrm{Hom}(V/Z, Z)$.) Notice that this gives another proof that $\mathrm{Gras}(k,n)$ is smooth and irreducible of dimension $(k+1)(n-k)$.

(b) Show that $\mathrm{Pic}(\mathrm{Gras}(k,n)) = \mathbb{Z}$, a generator being given by the class of a hyperplane section $H = \mathrm{Gras}(k,n) \setminus U$.

A.3. Linear Systems

In this section we will describe a correspondence between morphisms to projective space and families of effective divisors. The underlying idea is quite simple. To each embedding $X \hookrightarrow \mathbb{P}^n$ there corresponds its family of hyperplane sections, that is, the intersection of the image of X with all possible hyperplanes in $\mathbb{P}^n$. In this section we reverse this construction. For each family of effective divisors (a linear system) we will describe an associated rational map to a projective space. It is natural to ask under what conditions this map is a morphism or an embedding. This leads to the notion of base points and ampleness. Finally, we introduce the powerful language of sheaves and bundles, which are often more convenient than divisors. We explain why Cartier divisor classes are the same as isomorphism classes of line bundles (or line sheaves) and provide a dictionary for translating the various notions from one language to the other. In this section, whenever the variety we work with is not smooth, the word divisor means a Cartier divisor, since this notion is somewhat better behaved. For simplicity, throughout this section we assume that the base field k is algebraically closed. (See Section A.2.2, especially Proposition A.2.2.10, for an explanation of why this suffices for most applications.)

A.3.1. Linear Systems and Maps

Recall that to each divisor D on a variety X we have associated the vector space

$$L(D) = \{f \in k(X) \mid D + \operatorname{div}(f) \geq 0\} \cup \{0\},$$

whose dimension is denoted by $\ell(D)$. (Note that 0 is included in $L(D)$ by convention, or one could say that it is in $L(D)$, because the zero function vanishes to arbitrarily high order along every irreducible divisor.) The set of effective divisors linearly equivalent to D is naturally parametrized by the projective space

$$\mathbb{P}\big(L(D)\big) \cong \mathbb{P}^{\ell(D)-1}.$$

This parametrization is given by

$$\mathbb{P}\big(L(D)\big) \longrightarrow \{D' \mid D' \geq 0 \text{ and } D' \sim D\}, \qquad f \bmod k^* \longmapsto D + \operatorname{div}(f).$$

The following definition slightly generalizes this construction.

Definition. A *linear system* on a variety X is a set of effective divisors all linearly equivalent to a fixed divisor D and parametrized by a linear subvariety of $\mathbb{P}\big(L(D)\big) \cong \mathbb{P}^{\ell(D)-1}$. The *dimension* of the linear system is the dimension of the linear subvariety. (Some authors use the synonym *linear series*.)

Example A.3.1.1. The set of effective divisors linearly equivalent to D is a linear system called the *complete linear system of* D. It is denoted by $|D|$.

Another way to define a linear system L is to say that it is a subset of some $|D|$ such that $\{f \in k(X)^* \mid D + \operatorname{div}(f) \in L\} \cup \{0\}$ is a k-vector subspace of $k(X)$.

Example A.3.1.2. Let $D = 0$ be the zero divisor on $\mathbb{A}^n$. Then $L(D) = k[X_1, \dots, X_n]$, and $|D|$ is the set of all hypersurfaces in $\mathbb{A}^n$. Clearly, $L(D)$ and $|D|$ are infinite-dimensional.

Example A.3.1.3. Let d be a positive integer, and let H be the hyperplane $\{x_0 = 0\}$ in $\operatorname{Div}(\mathbb{P}^n)$. Then $L(dH) = \{F/x_0^d \mid F \in k[x_0, \dots, x_n]_d\}$, so $L(dH)$ has dimension $N = \binom{n+d}{n}$. The linear system $|dH|$ is the linear system of all hypersurfaces of degree d in $\mathbb{P}^n$; it has dimension $N - 1$.

These last two examples suggest that complete linear systems are likely to be more useful on projective varieties than they are on affine varieties. We will see below (A.3.2.7) that a linear system on a projective variety is always finite dimensional.

Example A.3.1.4. Let $X \hookrightarrow \mathbb{P}^n$ be a projective variety, let I_X be the homogeneous ideal of X, and let d be a positive integer. Each form F of degree d not in I_X cuts out an effective divisor $(F)_X$ on X; see (A.2.2.2). The collection of these divisors defines a linear system on X,

$$L_X(d) := \{(F)_X \mid F \in (k[x_0, \dots, x_n]/I_X)_d \setminus \{0\}\}.$$

As we will see below, this linear system determines the embedding $X \hookrightarrow \mathbb{P}^n$ up to a change of coordinates in $\mathbb{P}^n$.

Example A.3.1.5. Let $f : X \to Y$ be a morphism and let L be a linear system on Y such that $f(X)$ is not contained in any $D \in L$. Then the set of effective divisors $\{f^*D \mid D \in L\}$ is a linear system on X. Under some conditions we can even extend this to the case where f is only a rational map. Suppose that $f : X \to Y$ is a dominant rational map and that f is regular on $U := X \setminus Z$ with $\operatorname{codim}_X(Z) \geq 2$. If X is smooth, this last condition is automatic. We then observe that $(f|_U)^*D$ is a well-defined divisor on U. Further, the inclusion $U \subset X$ induces a natural map $\operatorname{Div}(X) \to \operatorname{Div}(U)$, which must be a bijection because of the codimension assumption. We may then define $f^*D \in \operatorname{Div}(X)$ as the unique divisor such that $(f^*D)|_U = (f|_U)^*D$. In this way we can pull linear systems back using rational maps.

We now describe the rational map associated to a finite-dimensional linear system.

Definition. Let L be a linear system of dimension n parametrized by a projective space $\mathbb{P}(V) \subset \mathbb{P}(L(D))$. Select a basis $f_0, \dots, f_n$ of $V \subset L(D)$. The *rational map associated to* L, denoted by ϕ_L, is the map

$$\begin{array}{rccl} \phi_L : & X & \longrightarrow & \mathbb{P}^n, \\ & x & \longmapsto & (f_0(x), \dots, f_n(x)). \end{array}$$

Remarks. (i) The map ϕ_L is clearly defined outside of the poles of the individual f_i's and the set of common zeros of the f_i's

(ii) The map ϕ_L depends on the choice of the basis, so it is well-defined only up to an automorphism of $\mathbb{P}^n$. It can be defined canonically with values in $\mathbb{P}(V)$.

(iii) If L is a linear system on X and D_0 is an effective divisor, then the set $\{D + D_0 \mid D \in L\}$ is also a linear system, and clearly the map it defines is the same as the map defined by L.

These remarks suggest the following definitions.

Definition. The set of *base points* of a linear system L is the intersection of the supports of all divisors in L. We will say that a linear system is *base point free* if this intersection is empty, and we will say that a divisor D is *base point free* if the complete linear system $|D|$ is base point free.

For a nonempty linear system L, it is easy to show that ϕ_L is regular outside of the base points of L (see Exercise A.3.5). However, the domain of ϕ_L need not be exactly the complement of the base points, simply because for any effective divisor E, the linear system $L' = E + L := \{D + E \mid D \in L\}$ defines the same rational map. This new linear system clearly has the support of E among its base points.

Definition. The *fixed component of a linear system* L is the largest divisor D_0 such that for all $D \in L$, we have $D \geq D_0$. If $D_0 = 0$, we say that the linear system has no fixed component.

We can now formulate the correspondence between rational maps and linear systems.

Theorem A.3.1.6. There *is a natural bijection between:*

(i) *Linear systems L of dimension n without fixed components.*

(ii) *Morphisms $\phi : X \to \mathbb{P}^n$ with image not contained in a hyperplane, up to projective automorphism. (That is, we identify two rational maps $\phi, \phi' : X \to \mathbb{P}^n$ if there is an automorphism $\alpha \in \mathrm{PGL}(n+1)$ such that $\phi' = \alpha \circ \phi$.)*

PROOF. The proof is not difficult. For details, see Mumford [4, Theorem 6.8] or Hartshorne [1, II.7.1 and II.7.8.1]. □

We close this section by giving some examples that illustrate the general theory.

Examples A.3.1.7. (a) Let $i : X \hookrightarrow \mathbb{P}^n$ be a projective variety not contained in any hyperplane. Notice that the linear system $L_X(1)$ (A.3.1.4) defines the embedding i. If $L \subset L_X(1)$, then the associated rational map is the linear projection with center the intersection of the hyperplanes in L. See Exercise A.3.1 for an analysis of the base points of this linear projection.

(b) More generally, the map associated to the linear system $L_X(d)$ described in Example A.3.1.4 is essentially the embedding i of X composed with the d-uple embedding (A.1.2.6(a)). Precisely, consider the composition $\Phi_d \circ i : X \hookrightarrow \mathbb{P}^n \to \mathbb{P}^N$. The ideal of X will contain homogeneous forms of degree d when d is large, so the map associated to $L_X(d)$ is the same map, but with the image restricted to the smallest linear subvariety containing $\Phi_d \circ i(X)$.

(c) The Cremona transformation (A.1.2.6(g)) is defined by the linear system of conics passing through the points $(0,0,1)$, $(0,1,0)$, and $(1,0,0)$.

A.3.2. Ampleness and the Enriques–Severi–Zariski Lemma

In this section we describe methods for determining whether a linear system provides an embedding.

Definition. A linear system L on a projective variety X is *very ample* if the associated rational map $\phi_L : X \to \mathbb{P}^n$ is an embedding, that is, ϕ_L is a morphism that maps X isomorphically onto its image $\phi_L(X)$. A divisor D is said to be *very ample* if the complete linear system $|D|$ is very ample. A divisor D is said to be *ample* if some positive multiple of D is very ample.

Notice that very ample divisors are hyperplane sections for some embedding. Also, the linear systems $L_X(d)$ are clearly very ample. We will see (A.3.2.5) that in some sense, up to composition with a d-uple embedding, all embeddings are given by such a linear system. Recall that a morphism is an embedding if it is injective and its tangent map at each point is injective. This allows us to give the following criterion.

Theorem A.3.2.1. *A linear system L on a variety X is very ample if and only if it satisfies the following two conditions:*

(i) (Separation of points) *For any pair of points $x, y \in X$ there is a divisor $D \in L$ such that $x \in D$ and $y \notin D$.*

(ii) (Separation of tangent vectors) *For every nonzero tangent vector $t \in T_x(X)$ there is a divisor $D \in L$ such that $x \in D$ and $t \notin T_x(D)$.*

PROOF. See Hartshorne [1, II, Proposition 7.3 and Remark 7.8.2]. □

For curves this translates easily into the following useful criterion.

Corollary A.3.2.2. *Let D be a divisor on a curve C.*
(a) *The divisor D is base point free if and only if for all $P \in C$ we have $\ell(D-P) = \ell(D) - 1$.*
(b) *The divisor D is very ample if and only if for all points $P, Q \in C$ we have $\ell(D - P - Q) = \ell(D) - 2$. (Note that we allow $P = Q$, which corresponds to the separation of tangent vectors condition in Theorem A.3.2.1.)*

PROOF. See Hartshorne [1, IV, Proposition 3.1] or Exercise A.3.2. □

These criteria enable us to prove that the set of ample divisors generates the group of divisors, and thus also generates the Picard group.

Theorem A.3.2.3. *Every divisor can be written as the difference of two (very) ample divisors. More precisely, let D be an arbitrary divisor and let H be a very ample divisor.*
(i) *There exists an $m \geq 0$ such that $D + mH$ is base point free.*
(ii) *If D is base point free, then $D + H$ is very ample.*

PROOF. Clearly, we may assume that $X \hookrightarrow \mathbb{P}^n$ and that H is a hyperplane section. We first use the moving lemma (A.2.2.5) to find divisors $D_1, \ldots, D_r$, all linearly equivalent to D, such that $\bigcap \operatorname{supp}(D_i) = \emptyset$. We may write D_i as an effective divisor minus a divisor $\sum_j m_{ij} Y_{ij}$, where each Y_{ij} is an irreducible subvariety of codimension 1 in X defined, say, by a set of forms F_{ijk}, and each m_{ij} is greater than 0. We select an integer

$$d \geq \max_i \left\{ \sum_{j,k} m_{ij} \deg F_{ijk} \right\}$$

and proceed to show that $D + dH$ is base point free.

Let $x \in X$, and choose some D_i such that $x \notin \operatorname{supp}(D_i)$. For each j there is then an index k_j with $F_{ijk_j}(x) \neq 0$. Also, let L be any linear form not vanishing at x, let $N = \sum_j m_{ij} \deg F_{ijk_j}$, and define

$$G = \left(\prod_j F_{ijk_j}^{m_{ij}} \right) L^{d-N}.$$

Note that $N \leq d$. Further, G is a form of degree d, so $(G)_X \sim dH$ and $D_i + (G)_X \sim D + dH$. On the other hand, we can compute

$$D_i + (G)_X \geq -\sum_j m_{ij} Y_{ij} + \sum_j m_{ij} (F_{ijk_j})_X + (d-N)(L)_X \geq 0.$$

Finally, we note that by construction, $x \notin \operatorname{supp}(D_i + (G)_X)$. This proves that x is not a base point of $D + dH$, which completes the proof of (i).

We assume now that D is base point free and prove that $D+H$ is very ample by verifying the criteria of Theorem A.3.2.1. Let x and y be distinct

points on X. Then we can find an effective divisor E, linearly equivalent to D, with $y \notin \operatorname{supp}(E)$, and a hyperplane section H_0 such that $x \in H_0$ but $y \notin H_0$. Clearly, $x \in \operatorname{supp}(E+H_0)$, whereas $y \notin \operatorname{supp}(E+H_0)$, which shows that the linear system separates points. Next, let $t \in T_x(X)$ be a tangent vector. Then we select an effective divisor E, linearly equivalent to D, such that $x \notin \operatorname{supp}(E)$, and we take a hyperplane section H_0 passing through x but not containing the line generated by t. Then $x \in \operatorname{supp}(E + H_0)$ and $t \notin T_x(E + H_0)$, which shows that the linear system separates tangent vectors. Hence the linear systems is very ample by Theorem A.3.2.1. □

Remark. Theorem A.3.2.3 can be used to give the following presentation of $\operatorname{Pic}(X)$. Define $\operatorname{Emb}(X)$ to be the free group generated by the embeddings of X into some $\mathbb{P}^n$. To any embedding ϕ we associate the divisor class $c_\phi = \phi^*(H) \in \operatorname{Pic}(X)$. By linearity we obtain a map, which we still call c, from $\operatorname{Emb}(X)$ to $\operatorname{Pic}(X)$. The previous theorem tells us that this map is a surjection. Let $H(X)$ be the kernel of the map $\operatorname{Emb}(X) \to \operatorname{Pic}(X)$. It is clear that the following three sorts of elements are in $H(X)$:

(i) Let $\alpha \in \operatorname{PGL}(n+1)$ and let $\phi : X \hookrightarrow \mathbb{P}^n$. Then $c_{\alpha\circ\phi} - c_\phi \in H(X)$.

(ii) Let $i : \mathbb{P}^n \hookrightarrow \mathbb{P}^{n+1}$ be a linear injection and let $\phi : X \hookrightarrow \mathbb{P}^n$. Then $c_{i\circ\phi} - c_\phi \in H(X)$

(iii) Let $\phi : X \hookrightarrow \mathbb{P}^n$ and $\psi : X \hookrightarrow \mathbb{P}^m$ be embeddings, let $S : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^N$ be the Segre embedding (A.1.2.6), let $\Delta : X \to X \times X$ be the diagonal embedding, and define a map by the composition

$$\phi \otimes \psi : X \xrightarrow{\Delta} X \times X \xrightarrow{\phi\times\psi} \mathbb{P}^n \times \mathbb{P}^m \xrightarrow{S} \mathbb{P}^N.$$

Then $c_{\phi\otimes\psi} - c_\phi - c_\psi \in H(X)$.

It can be shown that $H(X)$ is generated by these three types of elements.

We describe now the behavior of these notions under pullbacks.

Proposition A.3.2.4. (i) *Let $f : X \to Y$ be a morphism between two projective varieties. If D is a base point free divisor on Y, then f^*D is a base point free divisor on X.*

(ii) *Let $f : X \to Y$ be a finite morphism between two projective varieties. If D is an ample divisor on Y, then f^*D is an ample divisor on X.*

PROOF. The first part is very easy, but the second part is deeper. See Hartshorne [1, III, Exercise 5.7] for a proof. □

Notice that the pullback of a very ample divisor by a finite morphism need not be very ample. For example, consider the morphism ϕ described in (A.1.2.6(e)).

We now come to the proof that on a projective variety, the vector space $L(D)$ is finite-dimensional.

Theorem A.3.2.5. (Enriques–Severi–Zariski) *Let $X \hookrightarrow \mathbb{P}^n$ be a normal projective variety. There exists an integer $d_0 = d_0(X)$ such that for all integers $d \geq d_0$, the linear system $L_X(d)$ is a complete linear system. In other words, if D is an effective divisor on X that is linearly equivalent to d times a hyperplane section, then there is a homogeneous polynomial F of degree d such that $D = (F)_X$.*

PROOF. Let us denote by H a hyperplane section and suppose that D is an effective divisor on X linearly equivalent to dH. Writing H_i for the divisor cut out by $x_i = 0$ on X, we obtain functions f_i such that $\operatorname{div}(f_i) = D - dH_i$. Since $\operatorname{div}(f_i f_j^{-1}) = \operatorname{div}(x_i^{-d} x_j^d)$, we get relations $f_i f_j^{-1} = \lambda_{ij} x_i^{-d} x_j^d$ for certain constants $\lambda_{ij} \neq 0$. Multiplying the f_i's by some constant, we may assume that all λ_{ij} are equal to 1. Let $A = k[x_0, \ldots, x_n]/I_X$ be the homogeneous coordinate ring of X, let $\operatorname{Frac}(A)$ be its fraction field, and let A_M denote the homogeneous piece of A of degree M. We observe that the function $F = f_i x_i^d$ is independent of i and sits in $\operatorname{Frac}(A)$. Furthermore, the function f_i has no pole outside H_i; hence it lies in the integral closure of the affine ring of $X \smallsetminus H_i$, hence in the ring itself, since it is integrally closed (the variety X is normal by hypothesis). So each f_i has the shape $F_i/x_i^{m_i}$ for some homogeneous form F_i of degree m_i. So taking M large enough, we see that $x_i^{M-d} F = x_i^M f_i = x_i^{M-m_i} F_i$ belongs to A_M. We now apply the following lemma from commutative algebra.

Lemma A.3.2.6. *Let I be a homogeneous ideal in $k[x_0, \ldots, x_n]$, and let $A = k[x_0, \ldots, x_n]/I$. There exists an integer $d_0 = d_0(I)$ such that for all $d \geq d_0$, all $N \geq 0$, and all $F \in \operatorname{Frac}(A)$,*

$$x_0^N F, x_1^N F, \ldots, x_n^N F \in A_{N+d} \Longrightarrow F \in A_d.$$

PROOF. See Mumford [4, Proposition 6.11 part 2]. □

Returning to the proof of Theorem A.3.2.5, we see that if our d is larger than the d_0 given by Lemma A.3.2.6, then F is in A_d and D is the divisor cut out by (a representative of) F. This is the desired conclusion. □

We now have all the tools to complete the promised proof of finiteness of $\ell(D)$ for projective varieties. This is a special case of much more general finiteness results (see, for example, Hartshorne [1, Theorem III.2]).

Corollary A.3.2.7. *Let D be a divisor on a projective variety. Then $\ell(D) = \dim L(D)$ is finite.*

PROOF. We first assume that the projective variety X is normal. Clearly, we may assume that $\ell(D) \geq 1$. This means that D is linearly equivalent to an effective divisor. The dimension $\ell(D)$ depends only on the linear equivalence class of D, so we may take D to be effective. We also fix an embedding

$X \hookrightarrow \mathbb{P}^n$ and choose a homogeneous polynomial $G \in k[x_0, \ldots, x_n]$ that vanishes on the support of D, but not on all of X. We choose a G whose degree is larger than the integer $d_0(X)$ described in Theorem A.3.2.5. Replacing G by some power G^m to account for the multiplicities of the components of D, we may assume that $(G)_X \geq D$. It follows that $L(D) \subset L\big((G)_X\big)$. But from the previous theorem (A.3.2.5), we know that $L\big((G)_X\big) = L_X(d)$ is of finite dimension. Hence the same is true of $L(D)$. Now, if X is not normal, consider its normalization $\nu : X' \to X$ (see Exercise A.1.15); the map ν^* provides an injection of $L(D)$ into $L(\nu^*(D))$, and the latter has finite dimension by the previous argument. □

A.3.3. Line Bundles and Sheaves

In this section we introduce the powerful and versatile language of sheaves and (vector) bundles. These two objects are used throughout modern mathematics and already occur in the very definition of schemes. We give here only the most basic definitions and explain how the previous two sections can be reformulated in this terminology. This turns out to be more than mere paraphrasing, since the use of sheaves and bundles simplifies constructions and proofs and provides valuable insights. The reader should be aware that for the arithmetic applications contained in the subsequent parts of this book we will need only the theory of 1-dimensional vector bundles (i.e., line bundles).

Sheaves are devices to describe the local behavior of objects and to describe how local information is glued together to form global objects. In order to motivate the definition, we look at a familiar example from topology. Consider the set of continuous functions from a topological space X into another topological space Y. More generally, for any open subset $U \subset X$ we can look at the set of continuous functions $U \to Y$, which we denote by $\mathcal{C}(U)$ or $\mathcal{C}(U, Y)$. If $V \subset U$, then the restriction of a continuous function to V is again continuous, and hence we get a restriction map $\operatorname{res}_V^U : \mathcal{C}(U) \to \mathcal{C}(V)$. These maps have the following obvious compatibility: If $W \subset V \subset U$, then $\operatorname{res}_W^U = \operatorname{res}_W^V \circ \operatorname{res}_V^U$. Next we observe that two functions that agree locally are equal. That is, if we cover U by open sets, $U = \bigcup_i U_i$, and if two functions $f, g \in \mathcal{C}(U)$ satisfy $\operatorname{res}_{U_i}^U(f) = \operatorname{res}_{U_i}^U(g)$ for all i, then $f = g$.

Conversely, if we know locally functions that "match on the overlap," then we can glue them together. In other words, suppose that we have a covering $U = \bigcup_i U_i$ as above; suppose that we are given functions $f_i \in \mathcal{C}(U_i)$; and suppose that for all pairs of indices i, j we have $\operatorname{res}_{U_i \cap U_j}^{U_j}(f_j) = \operatorname{res}_{U_i \cap U_j}^{U_i}(f_i)$. Then there exists a function $f \in \mathcal{C}(U)$ satisfying $\operatorname{res}_{U_i}^U(f) = f_i$.

This example serves as a guide to the definition of a sheaf. We will soon realize that, like Monsieur Jourdain, we have already used several sheaves without knowing that we were doing so.

Definition. Let X be a topological space. A *presheaf* $\mathcal{F}$ on X consists of the following data:

(i) For every open subset U in X, a set $\mathcal{F}(U)$.

(ii) For all open subsets $V \subset U \subset X$, a map $r_{U,V} : \mathcal{F}(U) \to \mathcal{F}(V)$ satisfying

$$r_{U,U} = id_{\mathcal{F}(U)} \qquad \text{and} \qquad r_{U,W} = r_{V,W} \circ r_{U,V}.$$

In many cases we may think of the maps $r_{U,V}$ as *restriction maps.* This is especially true if they happen to be injective. If the $\mathcal{F}(U)$'s have some additional structure, for example if the $\mathcal{F}(U)$'s are groups, rings, or modules over some ring, then we speak of a *presheaf of groups, rings,* or *modules.*

Definition. A *morphism of presheaves* $f : \mathcal{F}_1 \to \mathcal{F}_2$ is a collection of maps $f(U) : \mathcal{F}_1(U) \to \mathcal{F}_2(U)$ such that for every $V \subset U$, the maps $f(U)$ and $f(V)$ are compatible with restriction, $r^2_{U,V} \circ f(U) = f(V) \circ r^1_{U,V}$. If the $\mathcal{F}_i$'s are presheaves of groups (respectively rings, modules), then we insist that the $f(U)$'s should be group (respectively ring, module) homomorphisms.

A presheaf on X attaches a set to each open subset of X, and also assigns various restriction maps. A sheaf is a presheaf in which local data determines global properties. In other words, if $U = \bigcup_i U_i$ is an open covering of U, then $\mathcal{F}(U)$ should be completely determined by the $\mathcal{F}(U_i)$'s, the $\mathcal{F}(U_i \cap U_j)$'s, and the various restriction maps connecting them. We make this precise in the following definition.

Definition. Let X be a topological space. A *sheaf* $\mathcal{F}$ on X is a presheaf with the property that for every open subset $U \subset X$ and every open covering $U = \bigcup_i U_i$, the following two properties are true:

(1) Let x, y be elements of $\mathcal{F}(U)$ such that $r_{U,U_i}(x) = r_{U,U_i}(y)$ for all i. Then $x = y$.

(2) Let $x_i \in \mathcal{F}(U_i)$ be a collection of elements such that for every pair of indices i, j, we have $r_{U_i,U_i\cap U_j}(x_i) = r_{U_j,U_i\cap U_j}(x_j)$. Then there exists a (unique) $x \in \mathcal{F}(U)$ such that $r_{U,U_i}(x) = x_i$ for all i.

These properties can be paraphrased as follows:

- Elements are uniquely determined by their local behavior;
- Compatible local data can be patched together (in a unique way) to form a global element.

Examples A.3.3.1.

(a) The fundamental example in classical algebraic geometry is the sheaf of regular functions on a variety X equipped with the Zariski topology. Thus $\mathcal{O}_X$ is the sheaf defined by

$$\mathcal{O}_X(U) = \{\text{regular functions on } U\},$$

and $r_{U,V}$ is the natural restriction of a function from U to V. It is immediate that $\mathcal{O}_X$ is a sheaf of rings. This construction is so fundamental that from the point of view of schemes (see Section A.9), a variety is a pair $(X, \mathcal{O}_X)$.
(b) The sheaf of invertible functions $\mathcal{O}_X^*$ associates to an open set U the set of regular functions without zeros on U. It is a sheaf of groups. Notice that $\mathcal{O}_X^*(U)$ is exactly the group of units in the ring $\mathcal{O}_X(U)$, hence the notation.
(c) On a variety X, the sheaf of rational functions $\mathcal{K}_X$ attaches to each open set U the set of rational functions on U. It is a constant sheaf in the sense that all of the maps $r_{U,V}$ are isomorphisms.
(d) The sheaf of differential r-forms Ω_X^r on a variety X associates to an open set U the set of regular r-differentials on U.
(e) Our motivating example, the presheaf of continuous functions on a topological space, is a sheaf. Similarly, on a C^∞ manifold X we can define a sheaf of C^∞ functions by the rule

$$\mathcal{C}^\infty(U) = \{f : U \to \mathbb{R} \mid f \text{ is a } C^\infty \text{ function}\}.$$

In modern language, the different types of geometry (e.g., differential, analytic, algebraic) are defined by attaching a certain type of structure sheaf to a certain type of topological space.

There is an obvious way to form the direct sum and tensor product of two sheaves of modules:

$$(\mathcal{F} \oplus \mathcal{G})(U) = \mathcal{F}(U) \oplus \mathcal{G}(U) \qquad \text{and} \qquad (\mathcal{F} \otimes \mathcal{G})(U) = \mathcal{F}(U) \otimes \mathcal{G}(U).$$

Next we consider what happens when we look at a sheaf in an infinitesimal neighborhood of a point. This is the algebraic analogue of the germs of functions used in analysis.

Definition. The *stalk* of a sheaf $\mathcal{F}$ at a point $x \in X$, by $\mathcal{F}_x$, is the direct limit of the $\mathcal{F}(U)$'s over all open sets U containing x. Thus

$$\mathcal{F}_x = \varinjlim_{x \in U} \mathcal{F}(U),$$

where the limit is taken with respect to the restriction maps $r_{U,V}$. Intuitively, an element of the stalk $\mathcal{F}_x$ is an element $s \in \mathcal{F}(U)$ for some open set containing x, where we identify s and $s' \in \mathcal{F}(U')$ if s and s' have the same restriction to $U \cap U'$.

It is clear that the stalk of a sheaf of groups (respectively rings, modules) is a group (respectively ring, module). The elements of $\mathcal{F}_x$ are called *germs* at x. If $x \in U$, we get a map $\mathcal{F}(U) \to \mathcal{F}_x$. The image of $s \in \mathcal{F}(U)$ in $\mathcal{F}_x$ is called the germ of s at x.

Example. Let $\mathcal{O}_X$ be the sheaf of regular functions on a variety X as described in (A.3.3.1(a)). Then the stalk of $\mathcal{O}_X$ at x is just the local ring $\mathcal{O}_{x,X}$.

The stalks of a sheaf contain local information. The local-to-global nature of sheaves is illustrated by the fact that many global properties of sheaves can be checked on the stalks. For example, a morphism $\mathcal{F}_1 \to \mathcal{F}_2$ of sheaves is an isomorphism if and only if all of the maps on the stalks, $\mathcal{F}_{1,x} \to \mathcal{F}_{2,x}$, are isomorphisms. See Hartshorne [1, Chapter II, Proposition 1.1].

Definition. Let $\mathcal{F}$ be a sheaf on X. The set of *global sections* of $\mathcal{F}$ is the set $\mathcal{F}(X)$. This set is also frequently denoted by $\Gamma(X, \mathcal{F})$.

For example, if X is an affine variety with coordinate ring $R = k[X]$, then $\Gamma(X, \mathcal{O}_X) = R$ and $\Gamma(X, \mathcal{O}_X^*) = R^*$. However, if X is a projective variety, then $\Gamma(X, \mathcal{O}_X) = k$ and $\Gamma(X, \mathcal{O}_X^*) = k^*$.

Some of the most important sheaves in algebraic geometry have the property that the sets $\mathcal{F}(U)$ are naturally modules over the ring $\mathcal{O}_X(U)$. We formalize this idea in the following definition.

Definition. Let X be a variety. An *$\mathcal{O}_X$-module* is a sheaf $\mathcal{F}$ on X such that:

(1) For every $U \subset X$, $\mathcal{F}(U)$ is a module over the ring $\mathcal{O}_X(U)$.

(2) For every $V \subset U \subset X$, the map $r_{U,V} : \mathcal{F}(U) \to \mathcal{F}(V)$ is a homomorphism of modules. In other words, if $s_1, s_2 \in \mathcal{F}(U)$ and $f_1, f_2 \in \mathcal{O}_X(U)$, then

$$r_{U,V}(f_1 s_1 + f_2 s_2) = r_{U,V}(f_1) r_{U,V}(s_1) + r_{U,V}(f_2) r_{U,V}(s_2).$$

Note that there are two different restriction maps $r_{U,V}$ here, one for $\mathcal{F}$ and one for $\mathcal{O}_X$.

For example, the sheaves $\mathcal{K}_X$ and Ω_X^r are clearly $\mathcal{O}_X$-modules. Similarly, the direct sum $\mathcal{O}_X \oplus \cdots \oplus \mathcal{O}_X = \mathcal{O}_X^r$ is an $\mathcal{O}_X$-module called a *free $\mathcal{O}_X$-module of rank r*.

Definition. Let $\mathcal{F}$ be an $\mathcal{O}_X$-module on X. We say that $\mathcal{F}$ is *locally free* if each point in X has a neighborhood over which $\mathcal{F}$ is free. The *rank* of a locally free sheaf $\mathcal{F}$ is the integer r such that $\mathcal{F}(U) \cong \mathcal{O}_X(U)^r$ for all sufficiently small open sets U. A locally free sheaf of rank 1 is called an *invertible sheaf* (or sometimes a *line sheaf*).

For example, when X is smooth, the sheaf of r-forms Ω_X^r is a locally free sheaf. This is a reformulation of Proposition A.1.4.6. On the other hand, the sheaf of rational functions $\mathcal{K}_X$ is not locally free. The reason that locally free sheaves of rank 1 are called "invertible" is because they are the sheaves $\mathcal{F}$ for which there exists another sheaf $\mathcal{F}'$ such that $\mathcal{F} \otimes \mathcal{F}' \cong \mathcal{O}_X$. (See Exercise A.3.13.) Thus the set of invertible sheaves naturally form a group, using tensor product as the group law and $\mathcal{O}_X$ as the identity element.

Remark. We now reinterpret the notion of a Cartier divisor by associating an invertible sheaf to each Cartier divisor. Let $D = \{(U_i, f_i) \mid i \in I\}$ be a Cartier divisor. We define the sheaf $\mathcal{L}_D$ to be the subsheaf of $\mathcal{K}_X$ determined by the conditions

$$\mathcal{L}_D(U_i) = \frac{1}{f_i}\mathcal{O}_X(U_i) \qquad \text{for all } i \in I.$$

This determines $\mathcal{L}_D$, since the U_i's cover X. It is not hard to check that $\mathcal{L}_D$ is well-defined and is locally free of rank 1. It is also easy to see that up to isomorphism, $\mathcal{L}_D$ depends only on the linear equivalence class of D, and that

$$\mathcal{L}_{D+D'} = \mathcal{L}_D \otimes \mathcal{L}_{D'}.$$

The association $\mathrm{Cl}(D) \mapsto \mathcal{L}_D$ thus defines a homomorphism from $\mathrm{Pic}(X)$ to the group of invertible sheaves (modulo isomorphism), and one can show that this map is in fact an isomorphism. See Hartshorne [1, Proposition II.6.13] or Shafarevich [1, Theorem 3, Chapter VI.1.4].

It turns out that locally free sheaves (of finite rank) can also be described by a more geometric object. The basic idea is that a locally free sheaf on X corresponds to a family of vector spaces parametrized continuously by the points of X.

Definition. A *vector bundle of rank* r over a variety X is a variety E and a morphism $p : E \to X$ with the following two properties:
(1) Each fiber $E_x = p^{-1}\{x\}$ is a vector space of dimension r.
(2) The fibration p is locally trivial. This means that for each point $x \in X$ there is a neighborhood U containing x over which the fibration is trivial. In other words, if we write $E_U = p^{-1}(U)$, then there is an isomorphism ϕ_U from E_U to $U \times \mathbb{A}^r$ such that the following diagram is commutative:

$$\begin{array}{ccc} E_U & \xrightarrow{\phi_U} & U \times \mathbb{A}^r \\ & {}_p\searrow \quad \swarrow_{p_1} & \\ & U & \end{array}$$

Here p_1 is the projection on the first factor. The maps ϕ_U are called *local trivializations* of E. A vector bundle of rank 1 is called a *line bundle*.

We will also need the notion of morphisms between vector bundles. They are morphisms of varieties that respect the bundle structure.

Definition. Let $p : E \to X$ and $p' : E' \to X'$ be vector bundles. A *morphism of vector bundles* is a pair of morphisms $f : E \to E'$ and $\bar{f} : X \to X'$ such that $\bar{f} \circ p = p' \circ f$ and such that for every $x \in X$, the map $f_x : E_x \to E'_x$ is a linear transformation of vector spaces.

The *trivial bundle* of rank r over X is $X \times \mathbb{A}^r \to X$. The following example of a nontrivial bundle is fundamental in (projective) algebraic geometry.

Example A.3.3.2. Let $V = \mathbb{A}^{n+1}$, and consider $\mathbb{P}^n$ to be the set of lines of V through 0. Define a variety

$$E = \{(x, v) \in \mathbb{P}^n \times V \mid v \text{ lies on the line } x\}.$$

Then projection onto the first factor, $p : E \to \mathbb{P}^n$, gives E the structure of a line bundle. Indeed, the first condition is clear, and it is easy to check that the fibration p trivializes above each standard affine open subset. Thus if we let $U_j = \mathbb{P}^n \setminus \{X_j = 0\}$, then the trivialization is given explicitly by

$$U_j \times \mathbb{A}^1 \longrightarrow E_{U_j}, \qquad (x, \lambda) \longmapsto \left(x, \left(\frac{\lambda x_0}{x_j}, \frac{\lambda x_1}{x_j}, \dots, \frac{\lambda x_n}{x_j}\right)\right).$$

We will develop tools below that can be easily used to show that E is not trivial, that is, E is not isomorphic to $\mathbb{P}^n \times \mathbb{A}^1$.

Example A.3.3.3. (Tangent bundle) Let $X \subset \mathbb{A}^n$ be an affine variety with ideal $I_X = (f_1, \dots, f_m)$. We define a variety $T(X)$ by

$$T(X) = \left\{ (x, t) \in X \times \mathbb{A}^n \;\middle|\; \sum_{i=1}^{n} \frac{\partial f_j}{\partial x_i}(x) t_i = 0 \text{ for all } 1 \le j \le m \right\},$$

and we let $p : T(X) \to X$ be projection onto the first factor. If X is smooth, then $T(X)$ is a vector bundle of rank equal to the dimension of X called the *tangent bundle of* X. The bundle $T(X)$ will be trivial over any open set where some fixed minor of the Jacobian $(\partial f_j / \partial x_i)$ does not vanish. This construction can be generalized to arbitrary varieties by taking an affine covering and then gluing the pieces together. For example, the tangent bundle of an algebraic group is trivial (see Exercise A.3.12).

Definition. Let $p : E \to X$ be a vector bundle. A *section of* E is a morphism $s : X \to E$ such that $p \circ s = \mathrm{id}_X$. Similarly, a *rational section of* E is a rational map $s : X \dashrightarrow E$ such that $p \circ s = \mathrm{id}_X$.

The set of sections to a vector bundle clearly form a vector space, which we will denote by $\Gamma(X,E)$. If X is a projective variety, it is easy to see that $\Gamma(X, X \times \mathbb{A}^r) = \mathbb{A}^r$, since a projective variety has no nonconstant regular functions. On the other hand, if E is the line bundle of Example A.3.3.2, then one can check that $\Gamma(\mathbb{P}^n, E) = \{0\}$. The advantage of considering rational sections is that a vector bundle always has nontrivial (i.e., nonzero) rational sections.

The connection between locally free sheaves and vector bundles is now easy to describe. Let $p : E \to X$ be a vector bundle. We associate to each open set U the vector space of sections $\Gamma(U, E_U)$. Notice that $\Gamma(U, E_U)$ is actually an $\mathcal{O}_X(U)$-module (by the rule $(fs)(x) = f(x)s(x)$). It is easy to check that the association $U \to \Gamma(U, E_U)$ defines a locally free sheaf $\mathcal{L}_E$ whose rank is equal to the rank of the vector bundle E. In fact, the association $E \mapsto \mathcal{L}_E$ is a bijection between (isomorphism classes of) vector bundles of rank r and (isomorphism classes of) locally free sheaves of rank r. See Hartshorne [1, II, Exercise 5.18] or Shafarevich [1, Theorem 2, VI.1.3]. For this reason we will use these notions interchangeably.

Next we describe how to construct bundles by gluing locally trivial bundles. The basic observation is that a vector bundle E and local trivializations ϕ_U give isomorphisms

$$\begin{array}{ccccc} E_{U_i \cap U_j} & \xrightarrow{\phi_{U_i}} & (U_i \cap U_j) \times \mathbb{A}^r & \xleftarrow{\phi_{U_j}} & E_{U_i \cap U_j} \\ & {\scriptstyle p}\searrow & \downarrow & \swarrow{\scriptstyle p} & \\ & & U_i \cap U_j & & \end{array}$$

We thus obtain isomorphisms $\phi_{U_j} \circ \phi_{U_i}^{-1} : (U_i \cap U_j) \times \mathbb{A}^r \to (U_i \cap U_j) \times \mathbb{A}^r$ that must be of the shape $(x, v) \mapsto (x, g_{ji}(x)v)$. Here g_{ji} is an $r \times r$ matrix with entries in $\mathcal{O}(U_i \cap U_j)$. The g_{ij}'s are called *transition functions*. The following identities are immediate:

$$g_{ii} = id \qquad \text{and} \qquad g_{ij}g_{jk} = g_{ik} \quad \text{on } U_i \cap U_j \cap U_k.$$

The set of g_{ij}'s determines the vector bundle E. Conversely, any set of g_{ij} satisfying these identities can be used to construct a vector bundle by gluing together trivial bundles.

Using this construction, we can define the *dual* of a vector bundle E to be the bundle $\check{E}$ whose fibers are the dual vector spaces of the fibers of E. Similarly, we define the *tensor product* of two bundles E and E' to be the bundle $E \otimes E'$ whose fibers are the tensor products of the fibers of E and E'. We also define the *pullback* of a bundle $p : E \to X$ by a morphism $f : Y \to X$ to be the fibered product

$$f^*E = E \times_X Y = \{(y, e) \in Y \times E \mid f(y) = p(e)\}.$$

Notice that if E and E' are line bundles, then $\check{E}$, $E \otimes E'$, and f^*E are also line bundles.

We next describe how to associate a line bundle to a Cartier divisor. A Cartier divisor on X is represented by a set of pairs $\{(U_i, f_i)\}_{i\in I}$, where the U_i's form a covering X and $f_i f_j^{-1} \in \mathcal{O}(U_i \cap U_j)$. We glue the trivial line bundles $U_i \times \mathbb{A}^1 \to U_i$ via the isomorphisms

$$(U_i \cap U_j) \times \mathbb{A}^1 \longrightarrow (U_i \cap U_j) \times \mathbb{A}^1, \qquad (x, \lambda) \longmapsto \big(x, \lambda(f_i f_j^{-1})(x)\big).$$

This gives a line bundle on X. We further observe that replacing the f_i's by $f_i f$ does not affect the construction, so the isomorphism class of the resulting line bundle depends only on the linear equivalence class of D.

Notation. Let D be a Cartier divisor. The line bundle associated to D as described above will be denoted by $\mathcal{O}(D)$.

Theorem A.3.3.4. *The association $D \mapsto \mathcal{O}(D)$ induces a functorial isomorphism between the group of Cartier divisor classes and the group of isomorphism classes of line bundles on X. More precisely,*

$$\mathcal{O}(D + D') = \mathcal{O}(D) \otimes \mathcal{O}(D') \qquad \text{and} \qquad \mathcal{O}(-D) = \mathcal{O}(D)\check{}\,.$$

Further, $\mathcal{O}(f^*D) = f^*\mathcal{O}(D)$ for any morphism f of varieties.

PROOF (sketch). One actually shows a little more. To each rational section s of a line bundle E one can associate a divisor $\operatorname{div}(s)$ such that $E = \mathcal{O}(\operatorname{div}(s))$. Further, the set of divisors obtained by varying s is the linear equivalence class

$$\{\operatorname{div}(s) \mid s \in \Gamma(X, \mathcal{O}(D)) \smallsetminus 0\} = |D|.$$

The functorial formulas then follow easily. See Shafarevich [1, Theorem 3, VI.1.4] for more details. □

We see that in this language, the space of sections $\Gamma(X, \mathcal{O}(D))$ is in bijection with the functions in $L(D)$. A linear system on X is thus given by choosing a vector subspace of $\Gamma(X, E)$ for some line bundle E on X.

Example A.3.3.5. It is classical (après Serre) to denote by $\mathcal{O}_{\mathbb{P}^n}(1)$ or $\mathcal{O}(1)$ the line bundle associated to a hyperplane. It is easy to see that $\mathcal{O}(1)$ is the dual of the line bundle defined in Example A.3.3.2. The global sections of $\mathcal{O}(1)$ can be identified with linear forms,

$$\Gamma(\mathbb{P}^n, \mathcal{O}(1)) = kX_0 \oplus \cdots \oplus kX_n.$$

We let $\mathcal{O}(d)$ denote the line bundle obtained by tensoring $\mathcal{O}(1)$ with itself d times. The global sections of $\mathcal{O}(d)$ are the homogeneous polynomials of degree d,

$$\Gamma(\mathbb{P}^n, \mathcal{O}(d)) = \bigoplus_{i_1+\cdots+i_n=d} kX_1^{i_1}\cdots X_n^{i_n}.$$

Observe that there is a natural product on sections of line bundles. If $s \in \Gamma(X, E)$ and $s' \in \Gamma(X, E')$, then $(s \otimes s')(x) = s(x) \otimes s'(x) \in E_x \otimes E'_x$ defines a section of $E \otimes E'$. With this in mind, the following translation of the Enriques–Severi–Zariski theorem (A.3.2.5) is immediate.

Corollary A.3.3.6. *Let $X \hookrightarrow \mathbb{P}^n$ be a normal projective variety and let D be a hyperplane section. Then for all sufficiently large d, the restriction map $\Gamma(\mathbb{P}^n, \mathcal{O}(d)) \to \Gamma(X, \mathcal{O}(dD))$ is surjective. In other words, every section (of a suitable power) of $\mathcal{O}(D)$ is given by homogeneous polynomials.*

Exercises

A.3.1. Let X be a projective variety, and let L be a linear subsystem of $L_X(1)$ with associated map ϕ_L. Let B be the intersection of the hyperplanes in L.

(a) Prove that ϕ_L is induced by the linear projection with center B.

(b) Describe the base locus of ϕ_L.

A.3.2. Let D be a divisor on a curve C. This exercise asks you to prove (A.3.2.2).

(a) Let $P \in C$. Prove that $\ell(D) \le \ell(D-P)+1$. Prove that $\ell(D) = \ell(D-P)+1$ if and only if there is a function f in $L(D)$ with $\operatorname{ord}_P(f) = \operatorname{ord}_P(D)$. (*Hint.* Use the injection $L(D-P) \hookrightarrow L(D)$.)

(b) Prove that D is base point free if and only if $\ell(D-P) = \ell(D) - 1$ for every point $P \in C$.

(c) Prove that D is very ample if and only if the linear system $|D-P|$ is base point free for all $P \in C$.

(d) Prove that D is very ample if and only if $\ell(D-P-Q) = \ell(D) - 2$ for every pair of (not necessarily distinct) points $P, Q \in C$.

A.3.3. Let L be a linear system of dimension 2 contained in the complete linear system of divisors of degree 3 on $\mathbb{P}^1$. Show that the associated map ϕ_L is a morphism $\phi_L : \mathbb{P}^1 \to \mathbb{P}^2$. Show that ϕ_L maps $\mathbb{P}^1$ birationally onto its image $\phi_L(\mathbb{P}^1)$, but that the map is not an isomorphism. (*Hint.* Show that the image has a singular point.)

A.3.4. (a) Show that up to (linear) automorphism, any rational map

$$f : \mathbb{P}^n \dashrightarrow \mathbb{P}^m$$

is the composition of a d-uple embedding (A.1.2.6(a)), a linear projection (A.1.2.6(c)), and a linear injection $(x_0, \dots, x_r) \mapsto (x_0, \dots, x_r, 0, \dots, 0)$.

(b) If $n > m$, prove that there are no nonconstant morphisms from $\mathbb{P}^n$ to $\mathbb{P}^m$.

(c) Prove that every morphism $\mathbb{P}^m \to \mathbb{P}^m$ is given by $m+1$ homogeneous polynomials $(P_0, \dots, P_m)$, where $P_1, \dots, P_m$ have no nontrivial common zeros.

A.3.5. Let L be a linear system on a variety X, let B_L be the set of base points of L, and let ϕ_L be the associated map $\phi_L : X \dashrightarrow \mathbb{P}^n$. Prove that the restriction of ϕ_L to $X \smallsetminus B_L$ is a morphism.

A.3.6. (a) Let D be a very ample divisor on X, let $\phi : X \to \mathbb{P}^n$ be the associated embedding, and let Y be a subvariety of X. Prove that the projective degree of $\phi(Y)$ is equal to the degree of Y with respect to D. In particular, the projective degree of $\phi(X)$ is the intersection number $D^n = (D, D, \dots, D)$.
(b) Let D be a divisor on X that is effective and ample, and let $Y \subset X$ be a subvariety of dimension at least one. Show that the intersection $Y \cap D$ is nonempty.

A.3.7. Let X be a projective variety of dimension n.
(a) Show that if D is (very) ample and $D' \sim D$, then D' is also (very) ample.
(b) Show that if D is ample, then (i) $D^n > 0$. (ii) For every subvariety $Y \subset X$, we have $Y.D^{\dim(Y)} > 0$.
(c) The Nakai–Moishezon criterion states that if a divisor has the positivity properties described in part (b), then it is ample. Use the Nakai–Moishezon criterion to show that if D is ample and if D' is algebraically equivalent to D, then D' is ample. (For a further discussion of the Nakai–Moishezon criterion, see Hartshorne [1, V, Theorem 1.10] for the case of surfaces and Hartshorne [1, Appendix A, Theorem 5.1] for the general case.)
(d) Give an example where D is very ample and D' is algebraically equivalent to D, but D' is not very ample (*Hint.* On a smooth curve, two divisors are algebraically equivalent if and only if they have the same degree.)

A.3.8. (a) Let L_1 and L_2 be two nonempty linear systems, and let L be the linear system spanned by the two. Show that the sum map from $L_1 \times L_2$ to L is algebraic with finite fibers.
(b) Let D_1 and D_2 be two divisors on a variety. Show that if $\ell(D_1) \geq 1$ and $\ell(D_2) \geq 1$, then $\ell(D_1) + \ell(D_2) \leq \ell(D_1 + D_2) + 1$.
(c) If D_1 and D_2 are base point free, prove that $\ell(D_1 + D_2) \leq \ell(D_1)\ell(D_2)$. (*Hint.* Use the morphisms ϕ_{D_i} and the Segre map.)

A.3.9. Show that Cartier divisors on a variety X, as we have defined them, can naturally be identified with global sections of the quotient sheaf $\mathcal{K}_X^*/\mathcal{O}_X^*$. (*Warning.* If $\mathcal{G}$ is a subsheaf of $\mathcal{F}$, it is not true in general that the presheaf $U \to \mathcal{F}(U)/\mathcal{G}(U)$ defines a sheaf. See Exercise A.3.15 below for the precise definition of the *quotient sheaf* $\mathcal{F}/\mathcal{G}$.)

A.3.10. Prove that $\mathcal{O}_X(U) = \bigcap_{x\in U} \mathcal{O}_{x,X}$ (the intersection is over all points $x \in U$). This shows that one can reconstruct the sheaf $\mathcal{O}_X$ from knowledge of the stalks (local rings). Prove also that one may reconstruct the function field of X from the sheaf $\mathcal{O}_X$ by the formula $k(X) = \varinjlim_U \mathcal{O}_X(U)$ (the limit is over all nonempty open sets $U \subset X$).

A.3.11. Let D and E two very ample divisors on a variety X, let $x_0, \dots, x_m$ (respectively $y_0, \dots, y_n$) be a basis of $\Gamma(X, \mathcal{O}(D))$ (respectively a basis of $\Gamma(X, \mathcal{O}(E))$), and let d and e be positive integers.
(a) Let s be a section of $\mathcal{O}(dD - eE)$. Prove that $y_i^e s$ is a section of $\mathcal{O}(dD)$. Deduce that if d is large enough, then there exist homogeneous polynomials P_i of degree d in $x_0, \dots, x_m$ such that $y_i^e s = P_i(x_0, \dots, x_m)$.
(b) Conversely, suppose that we are given polynomials $P_0, \dots, P_n$, homogeneous of degree d such that $y_j^e P_i(x_0, \dots, x_m) - y_i^e P_j(x_0, \dots, x_m) = 0$. Show that these define a global section of $\mathcal{O}(dD - eE)$ as in (a).

A.3.12. Let G be an algebraic group, let $T(G)$ be the tangent bundle of G as described in (A.3.3.3), and let $g \in G$. Consider the translation map $R_g(x) = gx$ and its associated differential, which we denote by $t_g : T_e(G) \to T_g(G)$. Prove that the map

$$\phi : G \times T_e(G) \longrightarrow T(G), \qquad (g, Y) \longmapsto (g, t_g(Y)),$$

is an isomorphism of vector bundles, and deduce that the tangent bundle $T(G)$ is trivial. Prove further that the vector bundle of differential forms over G is also trivial.

A.3.13. Let X be a variety.
(a) Let $\mathcal{F}$ be a locally free sheaf of rank r on X, and let $\check{\mathcal{F}}$ be the dual of $\mathcal{F}$. Prove that $\mathcal{F} \otimes \check{\mathcal{F}}$ is isomorphic to to the free sheaf $\mathcal{O}_X^{r^2}$. In particular, if $\mathcal{F}$ is an invertible sheaf (i.e., $r = 1$), then $\mathcal{F} \otimes \check{\mathcal{F}} \cong \mathcal{O}_X$.
(b) If $\mathcal{F}$ is locally free and if there exists a sheaf $\mathcal{F}'$ such that $\mathcal{F} \otimes \mathcal{F}' \cong \mathcal{O}_X$, prove that $\mathcal{F}$ is of rank one.
(c) Let $\mathcal{F}$ be a sheaf on X, and let $f : X \to Y$ be a morphism of varieties. Prove that the formula

$$(f_*\mathcal{F})(U) = \mathcal{F}(f^{-1}(U))$$

defines a sheaf $f_*\mathcal{F}$ on Y. If $\mathcal{F}$ is locally free (of rank r), is it necessarily true that $f_*\mathcal{F}$ is locally free (of rank r)?

A.3.14. Let $p : E \to X$ be a vector bundle defined via an open covering U_i and transition functions g_{ij}.
(a) Give a description of the dual vector bundle of E in terms of an open covering and transition functions. Give a similar description of f^*E, where $f : Y \to X$ is a morphism of varieties.
(b) If E has rank r, prove that $\check{E} \otimes E$ is a trivial bundle of rank r^2.

A.3.15. (Kernel, image, and quotient of sheaves)
(a) Let $\phi : \mathcal{F} \to \mathcal{G}$ be a morphism of sheaves. Prove that each of the maps

$$\begin{aligned} U &\longmapsto \ker(\phi(U) : \mathcal{F}(U) \to \mathcal{G}(U)), \\ U &\longmapsto \mathrm{Image}(\phi(U) : \mathcal{F}(U) \to \mathcal{G}(U)), \\ U &\longmapsto \mathcal{G}(U)/\phi(\mathcal{F}(U)), \end{aligned}$$

defines a presheaf on X, but that in general only the first one defines a sheaf.
(b) It is possible to attach to every presheaf $\mathcal{F}$ a "smallest" sheaf $\overline{\mathcal{F}}$ containing $\mathcal{F}$. One can describe $\overline{\mathcal{F}}$ by universal mapping properties, but the following description provides a concrete construction. We define $\overline{\mathcal{F}}(U)$ to be the set of functions $f : U \to \bigcup_{x \in U} \mathcal{F}_x$ such that for each $x \in U$, $f(x) \in \mathcal{F}_x$, and further such that there is a neighborhood V of x contained in U and a $g \in \mathcal{F}(V)$ such that for all $y \in V$, the germ g_y of g at y is equal to $f(y)$.

Using this description, prove that $\overline{\mathcal{F}}$ is a sheaf, that there is a natural inclusion $\mathcal{F} \to \overline{\mathcal{F}}$, and that $\mathcal{F}$ and $\overline{\mathcal{F}}$ have the same stalk at every point of X.

In particular, if $\mathcal{F}$ is already a sheaf, then $\overline{\mathcal{F}} = \mathcal{F}$. (See [Hartshorne 1, II.1.2] for further details.)

(c) Continuing with the notation from (a), the *image sheaf* $\operatorname{Image}(\phi)$ is the sheaf associated as in (b) to the presheaf $U \mapsto \operatorname{Image}(\phi(U))$. Similarly, the *quotient* (or *cokernel*) *sheaf* $\mathcal{G}/\phi(\mathcal{F})$ is the sheaf associated to the presheaf $U \mapsto \mathcal{G}(U)/\phi(\mathcal{F}(U))$. As an example, see Exercise A.3.9 for an interpretation of the quotient sheaf $\mathcal{K}_X^*/\mathcal{O}_X^*$.

A.4. Algebraic Curves

This section features algebraic curves, the heroes of this book. The most naive examples are affine plane curves given by $P(x, y) = 0$. A first guess would be that the higher the degree of the polynomial P, the more complicated the curve. This is not quite true. There is a subtler invariant called the genus, which is a much better measure of the complexity of the curve. The genus, usually denoted by g, is a nonnegative integer. If the curve C is projective and nonsingular and defined over $\mathbb{C}$, then its genus is the number of holes (or handles) in the Riemann surface $C(\mathbb{C})$.

A curve is a variety of dimension one, so its field of rational functions has transcendence degree one. It is natural to consider two curves equivalent if they have isomorphic function fields, since there will then be a birational isomorphism between them. We explain in Section A.4.1 that every curve is birational to a plane curve with only mild singularities, and that each equivalence class of curves contains exactly one smooth projective curve. We then focus our attention on these smooth models in Section A.4.2. The Riemann–Roch theorem is a basic tool that counts the dimension of linear systems and embeddings. It provides a convenient abstract definition of the genus and will be the basis of much of our subsequent work. We then display the basic trichotomy of curves, dividing our study into curves of genus 0, curves of genus 1, and curves of genus greater than or equal to 2. The curves of genus 0 are easy to analyze, at least over an algebraically closed field, since such curves are isomorphic to $\mathbb{P}^1$. Curves of genus 1 are more interesting, and we will show that they can be given the structure of an algebraic group (once a point has been selected as origin). Finally, we will discuss briefly the geometry of curves of higher genus. However, many of the deeper properties of curves of higher genus are best understood in terms of their associated Jacobian variety. We will discuss Jacobian varieties later, in Sections A.6 and A.8.

At the end of this section we include a very short subsection on algebraic surfaces, mainly aimed at treating the example of the product of

a curve with itself. This example will be vital in the proof of Mordell's conjecture.

We close this introduction with a brief table that exhibits the geometric and arithmetic trichotomies of curves. One of the most striking paradigms of modern Diophantine geometry is that the underlying geometry of a variety should determine the qualitative arithmetic properties of the variety. This idea has been fully realized in the case of curves. It is the goal of this book to explain and prove this realization as shown especially in the last column of the following table.

The Trichotomy of Curves C a smooth projective curve defined over a number field K				
Algebraic Geometry		Complex Geometry		Arithmetic
genus g of C	canonical divisor on C	universal cover of $C(\mathbb{C})$	constant curvature on $C(\mathbb{C})$	$C(K)$ (if $C(K)$ is not empty)
$=0$	$-K_C$ ample	$\mathbb{P}^1(\mathbb{C})$	$\kappa > 0$	$\mathbb{P}^1(K)$
$=1$	$K_C = 0$	$\mathbb{C}$	$\kappa = 0$	a finitely generated group
≥ 2	K_C ample	$\{\lvert z\rvert < 1\}$	$\kappa < 0$	finite

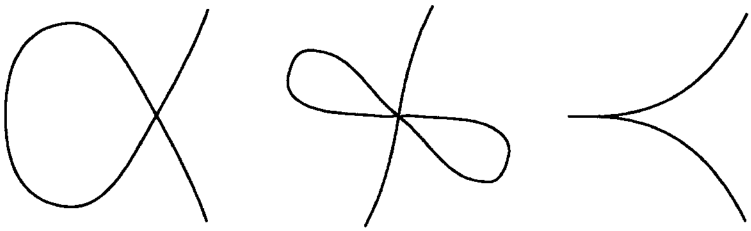
A.4.1. Birational Models of Curves

A curve C is a variety of dimension one, so its function field $k(C)$ is of transcendence degree one. It follows that $k(C)$ is algebraic over any subfield $k(x)$ generated by a nonconstant function $x \in k(C)$. Hence we may write $k(C) = k(x, y)$, where x and y are nonconstant functions on C satisfying an algebraic relation $P(x, y) = 0$. Let $C_0 \subset \mathbb{A}^2$ denote the affine plane curve defined by P, and let $C_1 \subset \mathbb{P}^2$ be the projective plane curve defined by the homogenized polynomial $Z^{\deg P} P(X/Z, Y/Z)$. Clearly, C is birational to both C_0 and C_1. Any curve birational to C is called a *model of* C (or of the function field $k(C)$), so we can say that every curve has a plane affine model and a plane projective model. It will soon be clear that these models cannot always be smooth (see the remarks after Theorem A.4.2.6), so we must allow singular points, but we should look for the mildest possible singularities. The next definition describes one sort of mild singularity.

Definition. An *ordinary singularity* is a singularity whose tangent cone is composed of distinct lines. The *multiplicity* of an ordinary singularity is the number of lines in its tangent cone.

For example, the point $P = (0, 0)$ is an ordinary singularity on the curve defined by $y^2 = x^3 + x^2$ (if $\operatorname{char}(k) \neq 2$), and more generally P is an ordinary singularity on the curve $y^n = x^{n+1} + x^n$ (if $\operatorname{char}(k) \nmid n$). On the other hand, $P = (0, 0)$ is not an ordinary singularity on the curve $y^2 = x^3$.

Notice that the multiplicity of an ordinary singularity P is just the number of distinct tangent directions at P. We have illustrated various singularities in Figure A.1.



$y^2 = x^3 + x^2$ Node $xy(x - y) = x^5 + y^5$ Triple point $y^2 = x^3$ Cusp

Ordinary and nonordinary singularities
Figure A.1

We now formally introduce some maps, already studied in (A.1.2.6(g)), that can be used to transform a complicated plane curve into a simpler one.

Definition. A *Cremona*, or *quadratic*, *transformation* is a birational involution from $\mathbb{P}^2$ to $\mathbb{P}^2$ that, after a linear change of variables on the domain and range, is defined by $Q(X, Y, Z) = (YZ, XZ, XY)$.

Theorem A.4.1.1. *An algebraic curve is birational to a plane projective curve with only ordinary singularities. More precisely, any plane curve can be transformed by a finite sequence of Cremona transformations into a plane curve with only ordinary singularities.*

PROOF. See Walker [1, III, Theorem 7.4] or Fulton [1, VII.4, Theorem 2]. One can, in fact, show that every curve is birational to a plane projective curve with only nodes as singularities (these are points with two distinct tangents), but this requires transformations more general than the Cremona transformations (see Hartshorne [1, IV, Corollary 3.6 and Theorem 3.10] for a proof). □

Theorem A.4.1.2. *A rational map from a smooth curve to a projective variety extends to a morphism defined on the whole curve.*

PROOF. This is a special case of Theorem A.1.4.4. □

Corollary A.4.1.3. *A birational morphism between two smooth projective curves is an isomorphism.*

PROOF. Clear from the previous theorem. □

We can now state the main result of this section.

Theorem A.4.1.4. *Any algebraic curve is birational to a unique (up to isomorphism) smooth projective curve.*

PROOF. See Walker [1, VI, Theorem 6.9], Fulton [1, VII.5, Theorem 3] or Hartshorne [1, I, Corollary 6.11]. The uniqueness follows immediately from Corollary A.4.1.3. Using more advanced commutative algebra, a quick proof of the existence can be given by constructing the normalization of the curve, which must be smooth. A more constructive or geometric proof consists in repeatedly blowing up the singular points and showing that this process eventually terminates. (See (A.1.2.6(f)) for the notion of blowing up.) For example, an ordinary singularity can be resolved by a single blowup (see Exercise A.4.1). □

A.4.2. Genus of a Curve and the Riemann–Roch Theorem

In view of Theorem A.4.1.4, we will concentrate on smooth projective curves. A divisor on such a curve C is simply a finite formal sum $D = \sum n_P P$, and we can define the *degree of* D to be $\deg(D) = \sum n_P$. We will denote a canonical divisor on C by K_C. Finally, we recall that

$$L(D) = \{f \in k(C) \mid (f) + D \geq 0\}$$

is a vector space of finite dimension $\ell(D)$. The Riemann–Roch theorem, which allows us to compute this dimension in most cases, is of inestimable value in the study of algebraic curves.

Theorem A.4.2.1. (Riemann–Roch theorem) *Let C be a smooth projective curve. There exists an integer $g \geq 0$ such that for all divisors $D \in \mathrm{Div}(C)$,*

$$\ell(D) - \ell(K_C - D) = \deg(D) - g + 1.$$

PROOF. See Serre [1, II.9, Théorème 3], Lang [4, I, Theorem 2.7], Hartshorne [1, IV, Theorem 1.3] or Fulton [1, VIII.6]. The "modern" proof is often divided into two parts. The first is a duality theorem expressing the left-hand side as an Euler–Poincaré characteristic. The second part is to calculate this Euler–Poincaré characteristic. □

The Riemann–Roch theorem is often stated and the proof given over an algebraically closed field, but using Proposition A.2.2.10, we see that it remains valid over any field of definition of C and D.

Definition. The integer g is called the *genus* of the smooth projective curve C. When C is not necessarily smooth or projective, its genus is defined to be the genus of the smooth projective curve that is birational to C (A.4.1.4).

It is tautological from this definition that the genus is a birational invariant. We next deduce several important corollaries from the Riemann–Roch theorem and devise various means of computing the genus of a curve.

Corollary A.4.2.2. *Let C be a smooth projective curve of genus g. Then*

$$\ell(K_C) = g \qquad \text{and} \qquad \deg(K_C) = 2g - 2.$$

PROOF. We first apply the Riemann–Roch theorem to the divisor $D = 0$ to get $1 - \ell(K_C) = -g + 1$. Note that $\ell(0) = 1$, since the only regular functions on a projective variety are the constant functions. Next we apply the theorem to $D = K_C$ to get $\ell(K_C) - 1 = \deg(K_C) - g + 1$. □

We have seen that there are no regular differentials on $\mathbb{P}^1$, so $\ell(K_C) = 0$. It follows from (A.4.2.2) that $\mathbb{P}^1$ has genus 0.

Corollary A.4.2.3. *Let C be a smooth projective curve of genus g and let $D \in \operatorname{Div}(C)$.*
(i) *If $\deg(D) < 0$, then $\ell(D) = 0$.*
(ii) *If $\deg(D) \geq 2g - 1$, then $\ell(D) = \deg(D) - g + 1$.*
(iii) (Clifford's theorem) *If $\ell(D) \neq 0$ and $\ell(K_C - D) \neq 0$, then we have $\ell(D) \leq \frac{1}{2}\deg(D) + 1$.*

PROOF. If $f \in L(D)$ is a nonzero function, then $D + (f)$ is effective, so $0 \leq \deg(D + (f)) = \deg(D)$. (Note that functions have degree 0.) Hence if $\deg(D) < 0$, then $L(D) = \{0\}$, so $\ell(D) = 0$. This proves (i), and then (ii) follows from (i), (A.4.2.2), and the Riemann–Roch theorem.

To prove (iii), we observe that the linear systems $|D|$ and $|K_C - D|$ are nonempty and that the addition map $|K_C - D| \times |D| \longrightarrow |K_C|$ is finite-to-one (Exercise A.3.8). Therefore, $\ell(K_C - D) - 1 + \ell(D) - 1 \leq \ell(K_C) - 1$. Combining this inequality with the Riemann–Roch theorem applied to D yields the desired result. □

Corollary A.4.2.4. *Let C be a smooth projective curve of genus g and let $D \in \operatorname{Div}(C)$.*
(i) *If $\deg(D) \geq 2g$, then D is base point free.*
(ii) *If $\deg(D) \geq 2g + 1$, then D is very ample.*
(iii) *D is ample if and only if $\deg(D) > 0$.*

PROOF. From the general considerations of Section A.3, especially Corollary A.3.2.2, we deduce that D is base point free if and only if $\ell(D - P) =$

$\ell(D) - 1$ for all $P \in C$. Similarly, we see that D is very ample if and only if $\ell(D - P - Q) = \ell(D) - 2$ for all $P, Q \in C$. Since $\ell(K_C - E) = 0$ when $\deg(E) > 2g - 2$, statements (i) and (ii) follow. Then (iii) is a simple consequence of (ii) and the fact that $\ell(D) = 0$ when $\deg(D) < 0$. □

A geometric version of (A.4.2.4(iii)) states that if C is a curve and if $U \subset C$ is an open subset of C with $U \neq C$, then U is affine. Indeed, it is true in general that the complement of an ample divisor is affine. We now describe a useful formula that can frequently be used to compute the genus of a curve.

Theorem A.4.2.5. (Riemann–Hurwitz formula) *Let C be a curve of genus g, let C' be a curve of genus g', and let $f : C \to C'$ be a finite separable map of degree $d \geq 1$. For each point $P \in C$, write e_P for the ramification index of f at P, and assume either that* $\operatorname{char}(k) = 0$ *or else that* $\operatorname{char}(k)$ *does not divide any of the e_P's. Then*

$$2g - 2 = d(2g' - 2) + \sum_{P \in C} (e_P - 1).$$

PROOF. From our analysis of differential forms (see Proposition A.2.2.8), we know that $f^*K_{C'} + \sum_{P \in C}(e_P - 1)P$ is a canonical divisor K_C on C. Taking degrees and using Corollary A.4.2.2 twice gives

$$\begin{aligned} 2g - 2 = \deg(K_C) &= \deg\Big(f^*K_{C'} + \sum_{P \in C} (e_P - 1)P\Big) \\ &= d\deg(K_{C'}) + \sum_{P \in C} (e_P - 1) = d(2g' - 2) + \sum_{P \in C} (e_P - 1). \end{aligned}$$

□

This formula gives a very convenient method to compute the genus of a curve, provided that we know, for example, one morphism from the curve to $\mathbb{P}^1$. It is also easy to deduce from it that the genus of a curve (in characteristic zero) is the number of handles of the associated Riemann surface. Another way to compute the genus is to write down the divisor of a differential form and then apply Corollary A.4.2.2. We illustrate this second method in the proof of the next theorem.

Theorem A.4.2.6. *Let C be a smooth projective plane curve of degree n. Then the genus g of C is given by the formula*

$$g = \frac{(n - 1)(n - 2)}{2}.$$

PROOF. We construct a (regular) differential form whose divisor has degree $n(n - 3)$. Then (A.4.2.2) implies that $n(n - 3) = \deg(K_C) = 2g - 2$, which gives the desired result.

Let $P(X,Y,Z) = 0$ be the homogeneous equation giving the curve. After a change of coordinates, we may assume that the line $Z = 0$ cuts the curve in n distinct points $Q_1, \dots, Q_n$, that none of these points lies on the line $Y = 0$, and that the function $v = Z/Y$ is a local parameter at each of the Q_i's. In other words, $\operatorname{ord}_{Q_i}(v) = 1$ for every i.

In the affine coordinates $(x,y) = (X/Z, Y/Z)$, the affine curve $U = C \smallsetminus \{Q_1, \dots, Q_n\}$ has equation $P(x,y,1) = 0$, and we may take $(u,v) = (X/Y, Z/Y)$ as coordinates near the Q_i's. We then define a differential form

$$\omega = \frac{dx}{P_Y(x,y,1)} = -\frac{dy}{P_X(x,y,1)} = \frac{v^{n-3}dv}{P_X(u,1,v)}.$$

(The subscripts indicate partial differentiation.) The fact that C is smooth means that $P(x,y,1)$, $P_Y(x,y,1)$, and $P_X(x,y,1)$ cannot all vanish, so ω has no poles on U. Further, either x or y is a local parameter at each point in U, so ω has no zeros on U, either. Thus $\operatorname{div}(\omega)_U = 0$. Finally, the last expression for ω shows that $\operatorname{ord}_{Q_i}(\omega) = n-3$ for every i, so $\operatorname{div}(\omega) = (n-3)\sum_{i=1}^{n}(Q_i)$. In particular, $\deg(\omega) = (n-3)n$, which completes the proof of the theorem.

□

Remarks. (i) Notice that the canonical divisor of a smooth plane curve of degree n is $n-3$ times a hyperplane section; hence the canonical divisor is very ample if $n \geq 4$. This implies (see Theorem A.4.5.1 below) that smooth plane curves of degree $n \geq 4$ are not hyperelliptic.

(ii) Continuing with the notation from the proof of Theorem A.4.2.6, the form $\omega_{ij} = x^i y^j \omega$ is easily seen to be regular on U if $i,j \geq 0$. It is also not hard to compute $\operatorname{ord}_Q(\omega_{ij})$ at the points $Q_1, \dots, Q_n$ and verify that ω_{ij} is regular at these points if and only if $i + j \leq n-3$. The set

$$\{x^i y^j \omega \mid i,j \geq 0,\ i+j \leq n-3\}$$

thus consists of $(n-1)(n-2)/2$ regular differential forms, and they are clearly linearly independent, so they provide a basis for the space $\Gamma(C, \Omega_C)$ of regular differential forms on C.

(iii) Lines and conics in $\mathbb{P}^2$ have genus 0, and in fact any curve of genus 0 defined over any field k is isomorphic over k to such a curve. (See Theorem A.4.3.1 for a precise statement.) Smooth cubics in $\mathbb{P}^2$ have genus 1, and every curve of genus 1 defined over k that also possesses a k-rational point is isomorphic to a plane cubic. (See Theorem A.4.4.1). On the other hand, it is clear that not every curve can be isomorphic to a smooth plane curve. In particular, the genus of a smooth plane curve is not arbitrary, since it must have the form $(n-1)(n-2)/2$. For example, a curve of genus 2, 4, 5, 7, 8 or 9 cannot have a smooth projective plane model.

When a plane curve has singularities, the formula for the genus must be modified.

Theorem A.4.2.7. *Let C be a projective plane curve of degree n with only ordinary singularities. Then its genus is given by the formula*

$$g = \frac{(n-1)(n-2)}{2} - \sum_{P \in S} \frac{m_P(m_P - 1)}{2},$$

where S is the set of singular points and m_P the multiplicity of C at P.

PROOF. See Walker [1, VI, Theorem 5.1] or Fulton [1 ,VIII.3, Proposition 5]. The formula of the theorem still holds for curves with arbitrary singularities, provided that we include in S the "infinitely near" points of a nonordinary singularity. See Hartshorne [1, V.3, page 392] for the definition of infinitely near points and a proof of the assertion. □

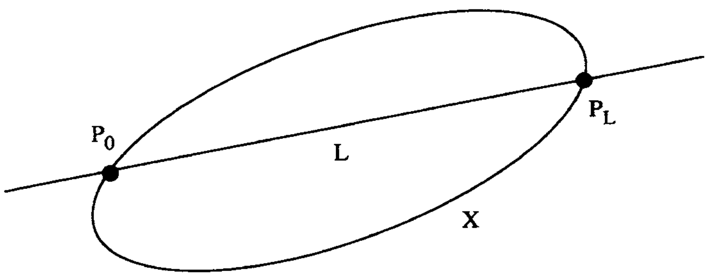
A.4.3. Curves of Genus 0

Let C be a (smooth projective) curve of genus 0 defined over a field k. Let K_C be a canonical divisor defined over k. Then $-K_C$ is a divisor of degree 2 (A.4.2.2), is defined over k, and is very ample (A.4.2.4(ii)). The Riemann–Roch theorem tells us that the dimension of the associated embedding is $\ell(-K_C) = 3$. Hence C can be embedded into $\mathbb{P}^2$ as a smooth curve X of degree 2 (i.e., as a conic) defined over k.

If $X(k) \neq \emptyset$, we can do better. Indeed let $P_0 \in X(k)$. We can identify $\mathbb{P}^1$ with the space of lines in $\mathbb{P}^2$ that go through P_0 and use this identification to define two maps as follows (see Figure A.2):

$$\phi : X \longrightarrow \mathbb{P}^1, \qquad P \longmapsto \begin{cases} \overleftrightarrow{PP_0} & \text{if } P \neq P_0, \\ \text{tangent line to } X \text{ at } P_0 & \text{if } P = P_0. \end{cases}$$

$$\psi : \mathbb{P}^1 \longrightarrow X, \qquad L \longrightarrow \text{the point } P_L \text{ such that } L \cap X = \{P_0, P_L\}.$$



The parametrization of a conic containing a rational point
Figure A.2

It is clear that these maps are (at least) rational maps whose composition is the identity. Now, either by an easy computation or by invoking Corollary A.4.1.3 we see that they are isomorphisms. We have thus proven most of the following result, and we leave the remaining easy bits for the reader.

Theorem A.4.3.1. *Let C be a smooth projective curve of genus 0 defined over a field k.*
(i) *The curve C is isomorphic over k to a conic in $\mathbb{P}^2$.*
(ii) *The curve C is isomorphic over k to $\mathbb{P}^1$ if and only if it possesses a k-rational point.*

Notice in particular that over an algebraically closed field, all curves of genus 0 are isomorphic to $\mathbb{P}^1$.

Definition. A curve is said to be *rational* if it is birational to the projective line. (*Warning.* Be sure you understand the two very different meanings of the word "rational" in the phrase "let P be a rational point on the rational curve C.")

Rational curves can be parametrized, and hence their set of rational points can be entirely described. Theorem A.4.3.1 says that a smooth projective curve is rational if it has genus 0 and possesses a rational point. Over a number field the existence of a rational point can be determined by the following important local-to-global criterion.

Theorem A.4.3.2. (Hasse principle) *A conic defined over a number field k has a k-rational point if and only if it has a rational point over all completions of k.*

PROOF. This is actually a special case of the Hasse principle, valid for all quadratic forms. See, for example, Serre [2, IV.3, Théorème 8]. □

We conclude by giving an easy geometric criterion for rationality.

Lemma A.4.3.3. *Let C be a smooth projective curve. Then the following are equivalent:*
(i) *C has genus 0.*
(ii) *There exists a point $P \in C$ such that $\ell(P) = 2$.*
(iii) *For every point $P \in C$ we have $\ell(P) = 2$.*

PROOF. Clearly, (iii) implies (ii), and applying (A.4.2.3(ii)) with $g = 0$ shows that (i) implies (iii). Finally, if (ii) holds, then the linear system associated to the divisor P gives a morphism $C \to \mathbb{P}L(P) = \mathbb{P}^1$ of degree one, which, since C and $\mathbb{P}^1$ are smooth curves, must be an isomorphism. This proves that (ii) implies (i). □

A.4.4. Curves of Genus 1

There are many books dedicated to unveiling the rich arithmetic structure of curves of genus 1. See, for example, Cassels [2], Knapp [1], Lang [5, 11], Silverman [1, 2], and the survey articles of Cassels [1] and Tate [2]. In this section we will be content to describe the group law on curves of genus 1, thereby providing our first examples of abelian and Jacobian varieties.

Let C be a curve of genus 1 defined over a field k. We know from (A.4.2.2) that any canonical divisor K_C has $\deg(K_C) = 0$ and $\ell(K_C) = 1$. Thus we can find an effective canonical divisor K_C with degree 0, which means that $K_C = 0$. In other words, the zero divisor is a canonical divisor. This means that there exists a regular differential form without zeros, a fact that will be explained when we show that C is an algebraic group.

Let D be a nonzero effective divisor on C. The Riemann–Roch theorem (with $g = 1$ and $K_C = 0$) tells us that $\ell(D) = \deg(D)$. Fix a point $P_0 \in C(k)$, and for each $n \geq 1$, consider the vector space

$$L_n = L\big(n(P_0)\big) \quad \text{whose dimension is} \quad \dim L_n = n.$$

Notice that these vector spaces are nested, $L_1 \subset L_2 \subset L_3 \subset \cdots$. Using our knowledge of the dimension of each L_n, we can find two functions $x, y \in k(C)$ such that

$$L_1 = k, \qquad L_2 = k \oplus kx, \qquad L_3 = k \oplus kx \oplus ky.$$

Notice that x has a pole of order 2 at P_0, that y has a pole of order 3 at P_0, and that x and y have no other poles. Using x and y, we can fill out L_4 and L_5,

$$L_4 = k \oplus kx \oplus ky \oplus kx^2, \qquad L_5 = k \oplus kx \oplus ky \oplus kx^2 \oplus kxy.$$

Notice that the functions $1, x, y, x^2, xy$ are linearly independent over k, since they each have different-order poles at P_0.

But when we look at L_6, we find that there are 7 functions that can be naturally constructed using x and y, namely

$$1,\ x,\ x^2,\ x^3,\ y,\ xy,\ y^2 \in L_6.$$

The vector space L_6 has dimension 6, so these functions satisfy a nontrivial k-linear relation. This gives part of the next result.

Theorem A.4.4.1. *Let C be a curve of genus 1 defined over a field k, and let $P_0 \in C(k)$. Then there exist constants $a_1, a_2, a_3, a_4, a_6 \in k$ such that C is isomorphic over k to the smooth plane cubic given by the equation*

$$E : Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3.$$

Under this isomorphism, the point P_0 is mapped to the inflection point $(X,Y,Z) = (0,1,0) \in E$.

If the characteristic of k is not 2 or 3, then by completing the square in y and the cube in x, one can find for E a curve given by an equation of the form

$$E : Y^2Z = X^3 + AXZ^2 + BZ^3, \qquad A, B \in k.$$

For a curve given in this form, the nonsingularity of E is equivalent to the nonvanishing of the discriminant $4A^3 + 27B^2 \neq 0$.

Definition. An *elliptic curve* is a pair (C, P_0), where C is a (smooth projective) curve of genus 1 and P_0 is a point on C. The elliptic curve is *defined over k* if the curve C is defined over k and also $P_0 \in C(k)$. Thus Theorem A.4.4.1 says that every elliptic curve is isomorphic to a smooth plane cubic with P_0 corresponding to an inflection point "at infinity." An equation of the form

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

or

$$Y^2Z = X^3 + AXZ^2 + BZ^3$$

is called a *Weierstrass equation* for E. Frequently, these equations are written in affine coordinates (i.e., by setting $Z = 1$), where it is understood that there is one additional point $P_0 = (0, 1, 0)$ at infinity.

PROOF (of Theorem A.4.4.1). We know by Corollary A.4.2.4 that the linear system associated to the vector space L_3 is a very ample linear system. This means that the rational map

$$\Phi : C \longrightarrow \mathbb{P}^2, \qquad P \longmapsto \big(1, x(P), y(P)\big),$$

extends to an isomorphism between C and its image $\Phi(C)$. In particular, $\Phi(C)$ is a smooth plane curve.

We have already observed above that x and y satisfy a relation of the form

$$ay^2 + bxy + cy = dx^3 + ex^2 + fx + g$$

for certain constants $a, b, c, d, e, f, g \in k$, not all zero. Further, x (respectively y) has a pole of order 2 (respectively 3) at P_0. Thus only the y^2 and x^3 terms have poles of order 6, so either the coefficients a and d are both nonzero, or they both vanish. But if $a = d = 0$, then every term has a different-order pole at P_0, so all of the other coefficients would have to vanish. Hence $ad \neq 0$. This allows us to replace (x, y) by (adx, ad^2y) and cancel a^3d^4, which gives an equation of the desired form.

Finally, if the characteristic of k is not 2 or 3, we can "complete the cube and the square" by making the linear transformation

$$X' = X + \frac{4a_2 + a_1^2}{12}Z, \qquad Y' = Y + \frac{a_1}{2}X + \frac{a_3}{2}Z.$$

This gives an equation $Y'^2Z' = X'^3 + AX'Z'^2 + BZ'^2$. We leave for the reader the easy task of verifying the nonsingularity condition. □

The labeling of the a_i coefficients of a Weierstrass equation is traditional and arises in the following way. Given any Weierstrass equation for E, we can make a change of coordinates $X' = u^2X$ and $Y' = u^3Y$ to obtain a new isomorphic equation for E. The a_i's of the new equation are then related to the old a_i's by the formula $a_i' = u^ia_i$. Thus the subscripts reflect the weights of the a_i's under change of coordinates.

For the remainder of this section we will assume that the characteristic of k is not 2 or 3, and we will work with an elliptic curve given by the affine Weierstrass equation

$$E : y^2 = x^3 + Ax + B.$$

However, everything we do can be formulated to carry over to the general case; see Silverman [1, III and Appendix A].

We begin by defining an involution $[-1]$ on E,

$$[-1] : E \longrightarrow E, \qquad (x, y) \longmapsto (x, -y).$$

Next we define a tangent and chord operation, which we will denote by "+." Let $P, Q \in E$. If P and Q are distinct, let L be the line through P and Q, while if $P = Q$, let L be the tangent line to E at P. The line L will meet E at a third point R, counting multiplicities, and then we set

$$P + Q = [-1]R.$$

(See Figure A.3.)

The next theorem justifies the notations $P+Q$ and $[-1]P$ by showing that with these operations, the points of E become a group. Furthermore, this group law is algebraic and intrinsic. That is, it is given by rational functions and does not depend on the particular equation or embedding of E. The group law depends only on the abstract curve E and the choice of the point P_0.

Theorem A.4.4.2. *Let E be a smooth projective cubic given by a Weierstrass equation*

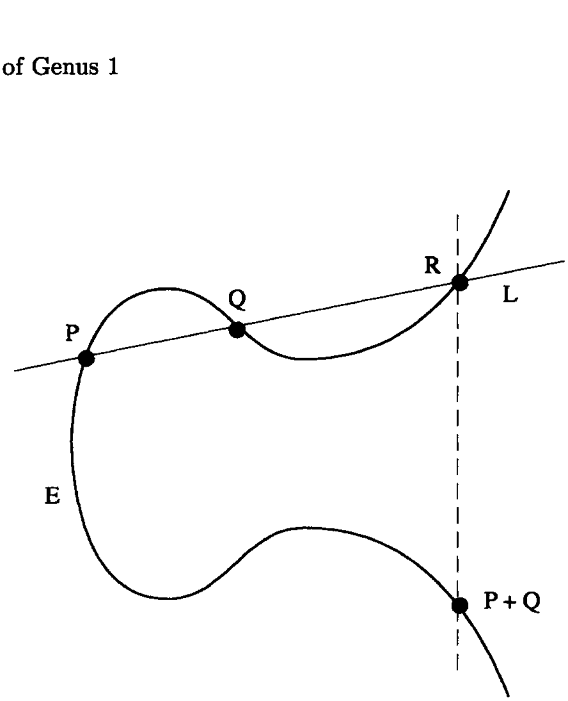
$$Y^2Z = X^3 + AXZ^2 + BZ^3.$$

Then the maps $(P, Q) \to P + Q$ and $P \to [-1]P$ defined above give E the structure of a commutative algebraic group with identity element $P_0 = (0, 1, 0)$. Furthermore, the map

$$\kappa : E \longrightarrow \operatorname{Pic}^0(E), \qquad P \longmapsto \text{divisor class of } (P) - (P_0),$$

is a group isomorphism.

PROOF. We start by explicitly computing the addition law, thereby verifying that the map "+" from $E \times E$ to E is algebraic. Let $P = (x_1, y_1)$



The addition law on a plane cubic curve E
Figure A.3

and $Q = (x_2, y_2)$ be two points of E distinct from the point at infinity. If $x_1 = x_2$ and $y_1 = -y_2$, then $P = [-1]Q$, and $P + Q$ is the point at infinity. Otherwise, let $y = \lambda x + \mu$ be the equation of the line L through P and Q. Thus

$$\lambda = \frac{y_1 - y_2}{x_1 - x_2} \quad \text{and} \quad \mu = \frac{y_2 x_1 - y_1 x_2}{x_1 - x_2} \qquad \text{if } P \neq Q,$$

$$\lambda = \frac{3x_1^2 + A}{2y_1} \quad \text{and} \quad \mu = \frac{-x_1^3 + Ax_1 + 2B}{2y_1} \qquad \text{if } P = Q.$$

To show that the two formulas for λ patch together to give an algebraic map on $E \times E$, we observe that a straightforward computation using the relation $y_i^2 = x_i^3 + Ax_i + B$ gives

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{x_1^2 + x_1 x_2 + x_2^2 + A}{y_1 + y_2}.$$

Setting $y_1 = y_2$ gives the desired equality. (Note that we are assuming that $y_1 \neq -y_2$.)

The intersection of L and E consists of the points satisfying the equations

$$y^2 = x^3 + Ax + B \qquad \text{and} \qquad y = \lambda x + \mu.$$

Eliminating y leads to a cubic in x, and we know that this cubic has x_1 and x_2 as two of its roots. Thus

$$x^3 - (\lambda x + \mu)^2 + Ax + B = (x - x_1)(x - x_2)(x - x_3) \qquad \text{for some } x_3.$$

Comparing coefficients, we find that $x_1 + x_2 + x_3 = \lambda^2$. This gives the value of x_3, and then substituting into $y = \lambda x + \mu$ gives the value of y_3. This proves that the intersection of the line L and the curve E consists of the three points (x_1, y_1), (x_2, y_2), and

$$(x_3, y_3) = (-x_1 - x_2 + \lambda^2, -\lambda x_1 - \lambda x_2 + \lambda^3 + \mu).$$

Finally, applying the involution $[-1]$, we obtain the explicit addition formula

$$(x_1, y_1) + (x_2, y_2) = (-x_1 - x_2 + \lambda^2, \lambda x_1 + \lambda x_2 - \lambda^3 - \mu),$$

where $\lambda, \mu \in k(x_1, y_1, x_2, y_2)$ are the rational functions given above.

This formula shows that the addition law $+ : E \times E \to E$ is a rational map, and our earlier remark shows that it is a morphism except possibly on the set where $x_1 - x_2 = y_1 + y_2 = 0$. But these are precisely the points that are mapped by addition to the point at infinity. We omit the verification that addition is again a morphism at these points (one needs to change coordinates or use a translation argument; see Silverman [1, III.3.6]). It is easy to see that the law is symmetric (abelian), that $P_0 := (0, 1, 0)$ is the identity element for the group, and that $[-1]$ provides an inverse. This justifies the notation.

It remains to show that addition on E is associative. This can be done by a direct calculation, using the explicit formulas, but the calculation is quite lengthy. We will avoid this by showing that the map κ is a bijection and that $\kappa(P + Q) = \kappa(P) + \kappa(Q)$, which clearly implies the rest of the theorem. (N.B. The symbol $+$ is being used here to represent two entirely different operations. When we write $\kappa(P + Q)$, the symbol $+$ means the addition law on the elliptic curve E, and so is given by the complicated, case-by-case formulas described above. When we write $\kappa(P) + \kappa(Q)$, we mean addition of divisor classes, which is a much easier operation.)

Suppose first that $P, Q \in E$ are distinct points with $\kappa(P) = \kappa(Q)$. This means that there is a function $f \in \bar{k}(C)$ with $\operatorname{div}(f) = (P) - (Q)$. But then the map $(f, 1) : E \to \mathbb{P}^1$ has degree 1, hence is an isomorphism, contradicting the fact that E has genus 1. This proves that κ is injective.

Next let $D \in \operatorname{Div}(E)$ be any divisor of degree 0. Then Riemann–Roch tells us that $\ell(D + (P_0)) = 1$, so there is an effective divisor (necessarily of degree 1) which is linearly equivalent to $D + (P_0)$. In other words, there is a point $P \in E$ with $(P) \sim D + (P_0)$, and hence $\kappa(P)$ is equal to the divisor class of D. This shows that κ is surjective.

It remains to show that κ satisfies $\kappa(P + Q) = \kappa(P) + \kappa(Q)$. Observe that the addition law can be characterized by the rule

$$P + Q + R = P_0 \quad \text{if and only if} \quad P,\ Q, \text{ and } R \text{ lie on a line.}$$

But P_0 is an inflection point, so the colinearity of P, Q, and R amounts to saying that $(P) + (Q) + (R) \sim 3(P_0)$. In other words,

$$P + Q + R = P_0 \quad \text{if and only if} \quad \kappa(P) + \kappa(Q) + \kappa(R) = 0.$$

Now everything is clear. □

Elliptic curves provide our first example of abelian varieties. When we define the Jacobian of a curve in Sections A.6 and A.8, we will rephrase the last statement of Theorem A.4.4.2 by saying that an elliptic curve is isomorphic to its own Jacobian. Note that it is quite remarkable that the abstract group $\text{Pic}^0(E)$ turns out to have a natural structure as an algebraic variety, just as it is remarkable that an abstract curve of genus 1 should have a natural group structure.

Finally, we would remiss if we did not point out that deciding whether a curve of genus 1 has any rational points can be a very difficult problem. In particular, if the ground field is a number field, the analogue of Theorem A.4.3.2 is false. There are curves C of genus 1 defined over a number field k such that $C(k_v) \neq \emptyset$ for every completion k_v of k, yet $C(k) = \emptyset$. A famous example, due to Selmer, is the curve

$$C : 3X^3 + 4Y^3 + 5Z^3 = 0.$$

(See Cassels [1, Appendix A], or Silverman [1, X.6.5], or Part C of this book.)

A.4.5. Curves of Genus at Least 2

We have already seen examples of curves of genus 0 and 1; in fact, we have seen all of them. It is considerably harder to describe all curves of genus $g \geq 2$, so we start modestly by giving some examples of such curves. One way to proceed is to try to describe a curve as a (finite) covering of $\mathbb{P}^1$, since every curve admits many such maps. We begin with the first nontrivial case.

Definition. A curve of genus $g \geq 2$ is called a *hyperelliptic curve* if it is a double covering of the projective line.

We will now describe these curves, where to simplify our discussion, we will work over an algebraically closed field k with $\text{char}(k) \neq 2$. The function field of a hyperelliptic curve C is a quadratic extension of $k(\mathbb{P}^1)$, hence has the shape $k(x, y)$, where $y^2 = F(x)$ for some polynomial $F(x) \in k[x]$. This equation gives an affine model for C. If the polynomial P has a double

root, say α, then we can replace y by $(x-\alpha)y$ and cancel $(x-\alpha)^2$. So we may assume that C is given by an affine equation

$$C : y^2 = F(x) \qquad \text{for some } F(x) \in k[x] \text{ with distinct roots.}$$

(See Exercise A.4.2 for details on this and on what follows.) The affine curve $y^2 = F(x)$ is smooth. We can obtain a complete smooth model for C by gluing this piece to the affine curve given by the equation

$$v^2 = F^*(u) := u^d F(u^{-1}),$$

where $d = \deg(F)$ if $\deg(F)$ is even, and $d = \deg(F) + 1$ if $\deg(F)$ is odd. The two affine pieces of C are glued together using the map $(u, v) = (x^{-1}, yx^{-d/2})$.

We can use the the Riemann–Hurwitz formula (A.4.2.5) to calculate the genus of C. Thus the map $C \to \mathbb{P}^1$ has degree 2 and is ramified at d points. (It is ramified at the points where $F(x) = 0$, and if the degree of F is odd, it is also ramified at the point at infinity.) The ramification index at each ramified point must be 2, so

$$2g - 2 = \deg(C \to \mathbb{P}^1)\big(2g(\mathbb{P}^1) - 2\big) + \sum_{P \in C} (e_P - 1) = -4 + d,$$

and we thus find that $g = (d/2) - 1$. In particular, C has genus 1 if and only if F has degree 3 or 4. Notice also that there exist hyperelliptic curves of every possible genus. (But you should be aware that many people use the term "hyperelliptic curve" only in reference to curves of genus at least 2.)

Rather than creating C by gluing together two affine curves, we can instead use the functions x and y to embed C into $\mathbb{P}^{g+1}$ via the map

$$C \longrightarrow \mathbb{P}^{g+1}, \qquad P \longmapsto \big(1, x(P), x(P)^2, \ldots, x(P)^g, y(P)\big).$$

Notice that the case $\deg(F) = 3$ is just the embedding of an elliptic curve into $\mathbb{P}^2$ described in (A.4.4.1).

We have already mentioned that curves of genus $g \geq 2$ are characterized by the fact that their canonical divisors are ample. The following theorem provides a further description of the canonical divisor.

Theorem A.4.5.1. *Let C be a smooth projective curve of genus g.*

(i) *The canonical divisor K_C is base point free if and only if $g \geq 1$.*
(ii) *The canonical divisor K_C is ample if and only if $g \geq 2$.*
(iii) *The canonical divisor K_C is very ample if and only if $g \geq 3$ and the curve is not hyperelliptic.*
(iv) *The bicanonical divisor $2K_C$ is very ample if and only if $g \geq 3$.*
(v) *The divisor $3K_C$ is very ample if and only if $g \geq 2$.*

PROOF. If K_C is not base point free, then there is a point $P \in C$ with $\ell(K_C - P) = \ell(K_C) = g$. Hence $\ell(P) = \ell(K_C - P) + 2 - g = 2$, and this

implies by Lemma A.4.3.3 that C is rational. This gives (i). Next, we have already seen that if C has genus 1, then $K_C = 0$. Riemann–Roch tells us that $\deg(K_C) = g$ and that K_C is ample if and only if $\deg(K_C) > 0$, which proves (ii). For the remaining parts of the theorem we may assume that $g \geq 2$.

Next we observe that any curve of genus 2 is hyperelliptic. More precisely, if C has genus 2, then Riemann–Roch says that $\ell(K_C) = \deg(K_C) = 2$, so the linear system $|K_C|$ gives a map of degree 2 from C to $\mathbb{P}^1$.

Corollary A.3.2.2 tells us that K_C is very ample if and only if for all points $P, Q \in C$, we have

$$\ell(K_C - P - Q) = \ell(K_C) - 2 = g - 2.$$

On the other hand, the Riemann–Roch theorem says that

$$\ell(P + Q) - \ell(K_C - P - Q) = \deg(P + Q) - g + 1 = 3 - g.$$

Combining these two equations, we find that

$$K_C \text{ is very ample} \iff \ell(P + Q) = 1 \text{ for all } P, Q \in C.$$

If C is hyperelliptic, say $C \to \mathbb{P}^1$, then the inverse image of any point in $\mathbb{P}^1$ consists of two points P, Q that satisfy $\ell(P+Q) = 2$. Thus K_C is not very ample on a hyperelliptic curve. Conversely, if K_C is not ample, then there are two points $P, Q \in C$ with $\ell(P + Q) = 2$, and hence the linear system $|P + Q|$ defines a map of degree 2 from C to $\mathbb{P}^1$. This completes the proof of (iii).

Next, if $g \geq 3$, then $\deg(2K_C) \geq 2g + 1$, so Corollary A.4.2.4 says that the divisor $2K_C$ is very ample. But if $g = 2$, then $\ell(2K_C) = 3$, so the linear system $|2K_C|$ maps C into $\mathbb{P}^2$, and we know that a plane curve of genus 2 cannot be smooth. This proves (iv). Finally, if $g \geq 2$, then $\deg(3K_C) \geq 2g + 1$, so (A.4.2.4) tells us that the divisor $3K_C$ is very ample. □

During the proof of the preceding theorem we showed that every curve of genus 2 is hyperelliptic. More precisely, if C has genus 2, then the linear system $|K_C|$ gives a map $C \to \mathbb{P}^1$ of degree 2. This gives a good description of all curves of genus 2. On the other hand, one can show that a "generic" curve of genus $g \geq 3$ is not hyperelliptic. For example, smooth plane quartics have genus 3 (Theorem A.4.2.6) and are not hyperelliptic, since a hyperplane section gives a canonical divisor (see Exercise A.4.3). In a certain sense that we will not make precise, the set of all (isomorphism classes of) curves of genus $g \geq 2$ is parametrized by a variety $\mathfrak{M}_g$ of dimension $3g - 3$, while the subset of hyperelliptic curves corresponds to a subvariety of $\mathfrak{M}_g$ of dimension $2g - 1$. Notice that these dimensions coincide for $g = 2$, in accordance with our observation that every curve of genus 2 is hyperelliptic. For an informal discussion of these varieties, called *moduli spaces*, see Mumford [3], and for a complete treatment see Mumford and Fogarty [1].

A.4.6. Algebraic Surfaces

This section is a very brief introduction to the geometry of smooth projective varieties of dimension two. Our main goal is to treat the case of the product of a curve with itself.

On a surface, divisors are formal sums of irreducible curves. (Note that the curves need not be smooth.) We can compute the intersection index of two such curves, or even the self-intersection of a single curve. The archetypical theorem of intersection theory is Bézout's theorem for the projective plane.

Theorem A.4.6.1. (Bézout's theorem) *Let C and D be curves on the surface $\mathbb{P}^2$ defined by irreducible equations of degree m and n. Then the intersection index of C and D is $C.D = mn$. In particular, if $C \neq D$, then the number of points of intersection counted with multiplicities is mn.*

PROOF. For a full proof, see Walker [1, IV.5, Theorem 4], Hartshorne [1, I, Corollary 7.8], Shafarevich [1, page 145] or Fulton [1, V.3]. We also observe that using the properties of the intersection index stated in Lemma A.2.3.1, the proof reduces to the case of two lines, where it is trivial. □

There is a remarkable connection between the genus of a curve and its self-intersection on a surface.

Theorem A.4.6.2. (Adjunction formula) *Let S be a smooth projective surface, let K_S be a canonical divisor on S, and let C be a smooth irreducible curve of genus g on S. Then*

$$C^2 + C.K_S = 2g - 2.$$

PROOF. See Hartshorne [1, V, Proposition 1.5] or Serre [1, IV.8, Lemme 2]. The formula is actually valid for a singular C, provided that we replace g by the arithmetic genus (see Serre [1, IV.8, Proposition 5]). □

Notice that (A.4.6.2) can be used to quickly rederive the formula for the genus of a smooth plane curve. Thus let $C \subset \mathbb{P}^2$ be a smooth plane curve of genus g and degree n, and let $H \subset \mathbb{P}^2$ be a line. Then $K_{\mathbb{P}^2} = -3H$, and $C \sim nH$, so the adjunction formula gives

$$2g - 2 = C^2 + C.K_{\mathbb{P}^2} = (nH)^2 + (nH).(-3H) = n^2 - 3n.$$

(Note that $H^2 = 1$, since any two lines are linearly equivalent, and two lines intersect in a single point.) One can similarly use (A.4.6.2) to compute the genus of curves lying on the quadric $\mathbb{P}^1 \times \mathbb{P}^1$; see Exercise A.4.4.

Just as for curves, one of the central results of the theory of surfaces is the Riemann–Roch theorem.

Theorem A.4.6.3. (Riemann–Roch for surfaces) *Let S be a smooth surface, and let K_S be a canonical divisor on S. There exists an integer $p_a(S)$ such that for any divisor $D \in \operatorname{Div}(S)$,*

$$\ell(D) - s(D) + \ell(K_S - D) = \frac{1}{2}D.(D - K_S) + 1 + p_a(S)$$

for some nonnegative integer $s(D)$.

PROOF. See Hartshorne [1, V, Theorem 1.6] or Serre [1, IV.8, Proposition 4]. The integer $p_a(S)$ is the *arithmetic genus* of S, and the integer $s(D)$ is historically called the superabundance. These quantities can be interpreted in terms of the dimensions of certain cohomology groups. For example, $s(D)$ is the dimension of $H^1(S, \mathcal{O}(D))$. □

Examples A.4.6.3.1. (i) The projective plane $\mathbb{P}^2$ has arithmetic genus $p_a(\mathbb{P}^2) = 0$.

(ii) Let C_1 and C_2 be smooth projective curves of genus g_1 and g_2 respectively. Then the arithmetic genus of the product $C_1 \times C_2$ is

$$p_a(C_1 \times C_2) = g_1 g_2 - g_1 - g_2.$$

Remark A.4.6.3.2. The integer $s(D)$ is often difficult to compute. When the characteristic of the ground field is zero, a useful tool is Kodaira's vanishing theorem, which states in our case that if D is ample, then $s(K_X + D) = 0$. The proof uses complex-analytic differential geometry and therefore does not extend to characteristic p. Indeed, there are known to be counterexamples in characteristic p.

Remark A.4.6.3.3. Let D be an ample divisor on a projective variety X. If X is a curve of genus g and if m is large enough ($m > (2g-2)/\deg(D)$ will suffice), then $\ell(mD) = m\deg(D) = 1 - g$. If X is a surface, one can show that $s(mD) = 0$ for m sufficiently large, and hence

$$\ell(mD) = m^2\frac{D \cdot D}{2} - m\frac{D \cdot K_X}{2} + 1 + p.$$

If X is an abelian variety, we will see (Theorem A.5.3.3) that

$$\ell(mD) = m^{\dim(X)}\frac{D^{\dim(X)}}{\dim(X)!}.$$

The general theorem of Riemann–Roch–Hirzebruch, combined with a vanishing theorem of Serre, shows that this remains approximately true in the general case, and we have

$$\ell(mD) = m^{\dim(X)}\frac{D^{\dim(X)}}{\dim(X)!} + O\left(m^{\dim(X)-1}\right).$$

We conclude this section by computing some interesting intersections on the product of a curve with itself.

Proposition A.4.6.4. *Let C be a smooth projective curve of genus g, fix a point P_0 on C, and let $S = C \times C$. Define divisors $D_1, D_2, \Delta \in \mathrm{Div}(S)$ by*

$$D_1 = C \times \{P_0\}, \qquad D_2 = \{P_0\} \times C, \qquad \Delta = \{(P,P) \in S \mid P \in C\}.$$

Notice that Δ is the diagonal of $C \times C$.
(i) $D_1^2 = D_2^2 = 0$.
(ii) $D_1.D_2 = \Delta.D_1 = \Delta.D_2 = 1$.
(iii) $\Delta^2 = 2 - 2g$.

PROOF. Part (i) follows by moving $\{P_0\}$ in its linear equivalence class so that the resulting divisors no longer intersect. Part (ii) is immediate from set theory, since the indicated pairs of divisors intersect transversally at the single point $(P_0, P_0) \in S$. Finally, to prove (iii), we apply the adjunction formula to Δ. As a curve, Δ is isomorphic to C, so its genus is g. On the other hand, the canonical divisor of a product is given by $K_S = K_C \times C + C \times K_C$ (see Exercise A.2.5(b)). It follows from the adjunction formula (A.4.6.2) that

$$\begin{aligned} 2g - 2 = \Delta^2 + \Delta.K_S &= \Delta^2 + \Delta.(K_C \times C) + \Delta.(C \times K_C) \\ &= \Delta^2 + 2\deg(K_C) = \Delta^2 + 4g - 4. \end{aligned}$$

Hence $\Delta^2 = 2 - 2g$. □

EXERCISES

A.4.1. Let C be a curve, let P be a point on C, and let C' be the blowup of the curve C at P.
(a) If P is a node, show that C' is smooth at the two points above P.
(b) More generally, suppose that P is an ordinary singularity with n distinct tangent directions. Prove that there are n distinct nonsingular points on C' lying above P

A.4.2. (Hyperelliptic curves) Recall that a smooth projective curve C of genus $g \geq 2$ is called *hyperelliptic* if there exists a double covering $\pi : C \to \mathbb{P}^1$. Let C be a hyperelliptic curve defined over a field k with $\mathrm{char}(k) \neq 2$.
(a) Show that C has an affine model U given by an equation of the form $y^2 = F(x)$, where $F(x)$ is a polynomial with distinct roots.
(b) Let $g = [(\deg F - 1)/2]$, and let $F^*(u) = u^{2g+2}F(u^{-1})$. Show that the equation $v^2 = F^*(u)$ also defines a smooth affine model U' of C.
(c) More precisely, show that there is an isomorphism

$$\{(x,y) \in U \mid x \neq 0\} \longrightarrow \{(u,v) \in U' \mid u \neq 0\}, \quad (x,y) \longmapsto (x^{-1}, yx^{-g-1}).$$

Prove that C is isomorphic to the curve obtained by using this map to glue U and U' together.

(d) Let U and U' be as in (c), and define a map

$$\phi : U \longrightarrow \mathbb{P}^{g+1}, \qquad (x, y) \longmapsto (1, x, x^2, \ldots, x^g, y).$$

Prove that ϕ is an embedding. Prove that the Zariski closure of $\phi(U)$ in $\mathbb{P}^{g+1}$ is smooth, and hence that it is isomorphic to C.

(e) Prove that the map $\pi : C \to \mathbb{P}^1$ is ramified at exactly $2g + 2$ points, and use the Riemann–Hurwitz formula to deduce that C has genus g. If C is given by the affine model $y^2 = F(x)$ with $\pi(x, y) = x$, identify the ramification points.

(f) Prove that the set $\{x^j dx/y \mid j = 0, 1, \ldots, g-1\}$ is a basis for the space of regular differential forms on C.

A.4.3. (Curves of genus 3)

(a) Show that on a smooth quartic curve in $\mathbb{P}^2$, a hyperplane section is a canonical divisor.

(b) If C is a smooth projective curve of genus 3, show that either C is hyperelliptic or else the canonical linear system $|K_C|$ embeds C as a plane quartic in $\mathbb{P}^2$. This gives a complete description of curves of genus 3.

A.4.4. Compute the genus of a smooth curve of bidegree (a, b) in $\mathbb{P}^1 \times \mathbb{P}^1$, and conclude that $\mathbb{P}^1 \times \mathbb{P}^1$ contains smooth curves of every genus. (This is in marked contrast to $\mathbb{P}^2$.)

A.4.5. (Curves of genus 4) Let C be a smooth projective curve of genus 4 that is not hyperelliptic.

(a) Show that the canonical linear system $|K_C|$ embeds C as a curve of degree 6 in $\mathbb{P}^3$.

(b) Prove that C lies on a quadric surface $Q = 0$ in $\mathbb{P}^3$. (*Hint.* Compare the dimension of $L(2K_C)$ to the dimension of the space of homogeneous polynomials of degree 2.)

(c) Prove that C also lies on an irreducible cubic surface $F = 0$.

(d) Conclude that C is the intersection of a quadric and a cubic.

(e) Conversely, prove that if a quadric surface and a cubic surface in $\mathbb{P}^3$ intersect in a smooth curve C, then C is a curve of genus 4 embedded via its canonical linear system. In particular, C not hyperelliptic.

A.4.6. Let $Q(x) \in k[x]$ be a polynomial of degree d with distinct roots, let C be a smooth projective curve with an affine model $y^n = Q(x)$, and assume that $\operatorname{char}(k)$ is 0 or does not divide n. Prove that the genus g of C satisfies

$$2g - 2 = nd - n - d - \gcd(n, d).$$

A.4.7. The purpose of this exercise is to prove the following theorem:

> **Theorem.** (Belyi) *Let C be a smooth projective curve defined over $\mathbb{C}$. Then C is defined over $\overline{\mathbb{Q}}$ if and only if there exists a finite map $C \to \mathbb{P}^1$ ramified only above the three points $\{0, 1, \infty\}$.*

(a) Assume that C is defined over $\bar{\mathbb{Q}}$. In order to prove that there is a finite map $C \to \mathbb{P}^1$ ramified only over $0, 1, \infty$, show that it suffices to prove the following statement: Let $S \subset \mathbb{P}^1(\bar{\mathbb{Q}})$ be finite set of algebraic points. Then there exists a finite map $f : \mathbb{P}^1 \to \mathbb{P}^1$ such that every ramification point of f and every point in S gets mapped by f into $\{0, 1, \infty\}$.
(b) Reduce the proof of the previous statement to the case where $S \subset \mathbb{P}^1(\mathbb{Q})$. (*Hint.* If $(a, 1) \in S$ with $[\mathbb{Q}(a) : \mathbb{Q}] = d \neq 1$, let F be the minimal polynomial of a and consider the map $x \mapsto F(x)$ from $\mathbb{P}^1$ to $\mathbb{P}^1$. Show that repeated application of this process will yield a map $f_1 : \mathbb{P}^1 \to \mathbb{P}^1$ that sends S and its ramification points into $\mathbb{P}^1(\mathbb{Q})$.)
(c) Let $c \in \mathbb{Q}^*$ and $a, b \in \mathbb{Z}$ with $a, b, a - b \neq 0$. Show that the map $x \mapsto cx^a(1 - x)^b$ from $\mathbb{P}^1$ to $\mathbb{P}^1$ is ramified only at the four points 0, 1, ∞, and $a/(a + b)$. Show that for an appropriate choice of c, the map sends these four ramification points into $\{0, 1, \infty\}$.
(d) Now use induction on the number of points in $S \subset \mathbb{P}^1(\mathbb{Q})$ to finish the proof of one direction of Belyi's theorem.
(e) Conversely, assuming that there is a finite map $C \to \mathbb{P}^1$ ramified only above $\{0, 1, \infty\}$, prove that C is defined over $\bar{\mathbb{Q}}$.

A.4.8. Let C be a smooth plane cubic curve defined over an algebraically closed field k with $\operatorname{char}(k) \neq 3$. Let $P_0 \in C(k)$ be an inflection point, and use P_0 as the identity element to give C a group structure. Prove that the points of order 3 in the group $C(k)$ are the other inflection points of C. Compute the number of such points, and use your result to describe the group structure of $\{P \in C(k) \mid 3P = P_0\}$.

A.4.9. Let C be a smooth cubic curve defined by $aX^3 + bY^3 + cZ^3 + dXYZ = 0$.
(a) Write down the condition on a, b, c, d for the curve to be smooth. (Notice in particular that the characteristic must be different from 3.)
(b) Let $P = (x, y, z) \in C$, and let L be the tangent line to C at P. Then $L \cap C$ consists of the point P with multiplicity 2 and a third point P'. Compute the coordinates of P' explicitly in terms of the coefficients of C and the coordinates of P.
(c) Assume now $k = \mathbb{Q}$ and that a, b, c, d are square-free integers with a, b, c distinct. Let $P = (x, y, z) \in C(\mathbb{Q})$ be a point with $x, y, z \in \mathbb{Z}$ and $\gcd(x, y, z) = 1$, and similarly write the point P' described in (b) as $P' = (x', y', z')$ with $x', y', z' \in \mathbb{Z}$ and $\gcd(x', y', z') = 1$. Prove that $|x'y'z'| > |xyz|$. Conclude that $C(\mathbb{Q})$ is either empty or infinite, and find examples of both instances.

A.4.10. Let k be a field with $\operatorname{char}(k) \neq 2, 3$, and let C be a smooth projective cubic curve given by the affine equation $y^2 = x^3 + Ax + B$. Prove directly that the space of regular differential forms on C has dimension 1 and that dx/y is a basis.

A.4.11. Let $f : C' \to C$ be a nonconstant (hence finite) separable morphism between smooth curves of genus g' and g, respectively.
(a) Prove that $g' \geq g$.
(b) Prove that $g' = g$ if and only if one of the following is true: (i) $g' = g = 0$. (ii) $g = 1$ and f is unramified. (iii) f is an isomorphism.
(c) Dropping the separability assumption, assume that C is defined over a field of characteristic $p > 0$, and let $F : C \to C^{(p)}$ be the Frobenius map

(see Exercise A.1.7). Then F is not an isomorphism (Exercise A.1.7(c)). Prove that C and $C^{(p)}$ have the same genus.

A.4.12. Let C be a hyperelliptic curve of genus $g \geq 2$ defined over k.

(a) Prove that the canonical map ϕ_{K_C} gives a map from C onto a rational curve C_0 of degree $g-1$ in $\mathbb{P}^{g-1}$. Further, show that the map ϕ_{K_C} and the rational curve C_0 can both be taken to be defined over k.

(b) If g is even, prove that C_0 has a k-rational point, and hence is isomorphic over k to $\mathbb{P}^1$.

(c) Give an example of a curve C of genus 3 defined over $\mathbb{Q}$ such that C is hyperelliptic over $\bar{\mathbb{Q}}$, but such that there does not exist a morphism $C \to \mathbb{P}^1$ of degree 2 defined over $\mathbb{Q}$.

A.4.13. Let C be a smooth curve in $\mathbb{P}^n$ of projective degree N not contained in any hyperplane.

(a) Prove that we can choose projective coordinates $x_0, \dots, x_n$ on $\mathbb{P}^n$ such that the following two conditions hold:

$$\text{(i)} \qquad C \cap \{x_i = x_j = 0\} = \emptyset \quad \text{for } 0 \leq i < j \leq n.$$
$$\text{(ii)} \qquad k(x_1/x_0, x_i/x_j) = k(C) \quad \text{for } 2 \leq i < j \leq n.$$

(b) Prove that there is a polynomial $G_{ij} \in k[X,Y]$ of total degree less than N^2 such that $G_{ij}(x_1/x_0, x_i/x_j) = 0$.

A.4.14. (Weierstrass points) Let C be a smooth projective curve of genus g. For each point $P \in C$, define a set of integers

$$G(P) = \{n \geq 1 \mid \ell(nP) = \ell((n-1)P)\}.$$

(a) Show that

$$\#G(P) = g \qquad \text{and} \qquad G(P) \subset \{1, 2, \dots, 2g-1\}.$$

Further, show that $\mathbb{N} \smallsetminus G(P)$ is a semigroup and that $n \in G(P)$ if and only if there is a regular differential form ω with $\operatorname{ord}_P(\omega) = n-1$. (*Hint.* Use the Riemann–Roch theorem.)

(b) Show that the following conditions on P are equivalent:

(i) $G(P) \neq \{1, 2, \dots, g\}$.

(ii) $\ell(gP) \geq 2$ (i.e., there is a nonconstant function f on C with a pole at P of order at most g and with no other poles).

(iii) There is a regular differential form ω on C with $\operatorname{ord}_P(\omega) \geq g$.

A point P satisfying these conditions is called a *Weierstrass point.* Show that a curve of genus 0 or 1 has no Weierstrass points.

(c) Define the Weierstrass weight of P to be

$$w(P) = \sum_{n \in G(P)} n - \frac{g(g+1)}{2}.$$

Prove the following properties of the Weierstrass weight.

(i) $w(P) \geq 0$.

(ii) $w(P) > 0$ if and only if P is a Weierstrass point.

(iii) $w(P) \le g(g-1)/2$.

(*Hint.* Show that the largest possible $w(P)$ is obtained when $G(P) = \{1, 3, \ldots, 2g-1\}$.)

(d) Let x be a local coordinate and let $\omega_1, \ldots, \omega_g$ be a basis of regular differential 1-forms. Write each ω_i as $\omega_i = f_i\,dx$ and consider the Wronskian determinant

$$W(x) = \det\left(\frac{d^{j-1}f_i(x)}{dx^{j-1}}\right)_{1\le i,j\le g}.$$

We define a differential form $\tilde{\omega}$ of weight $g(g+1)/2$ on C (i.e., a global section of $\Omega_1^{\otimes g(g+1)/2}$) by the formula $\tilde{\omega} = W(x)(dx)^{g(g+1)/2}$. Show that $\tilde{\omega}$ is independent (up to a scalar) of the choice of the local coordinate x and differentials $\omega_1, \ldots, \omega_g$. Prove that the divisor of $\tilde{\omega}$ is

$$\operatorname{div}(\tilde{\omega}) = \sum_{P\in C} w(P)P.$$

Deduce that there are only finitely many Weierstrass points on C.

(e) If C has genus $g \ge 2$, prove that

$$2g+2 \le \#\{\text{Weierstrass points of } C\} \le g^3 - g.$$

Show that the lower bound is possible by proving that a hyperelliptic curve has exactly $2g+2$ Weierstrass points. (*Hint.* They are the ramification points of the double cover $C \to \mathbb{P}^1$.)

A.4.15. Let C be a smooth projective curve C of genus g, and let $\operatorname{Aut}(C)$ denote the group of automorphisms of C.

(a) Prove that $\operatorname{Aut}(\mathbb{P}^1) = \operatorname{PGL}(2)$.

(b) Suppose now that $g = 1$, fix a point $P_0 \in C$, and use P_0 as usual to define a group law on C. For each point $Q \in C$ we can define a translation-by-Q map

$$t_Q : C \longrightarrow C, \qquad t_Q(P) = P + Q.$$

Prove that the map $C \to \operatorname{Aut}(C)$, $Q \to t_Q$, is an injective homomorphism.

Define a subgroup $G = \{\alpha \in \operatorname{Aut}(C) \mid \alpha(P_0) = P_0\}$, and define a map $\psi : \operatorname{Aut}(C) \to G$ by $\psi(\alpha)(P) = \alpha(P) - \alpha(P_0)$. Prove that the following sequence is exact and that it can be split:

$$0 \longrightarrow C \xrightarrow{\;t\;} \operatorname{Aut}(C) \xrightarrow{\;\psi\;} G \longrightarrow 0.$$

In other words, prove that $\operatorname{Aut}(C)$ is the semidirect product $C \rtimes G$ via the obvious action of G on C. Finally, prove that the group G is finite. (See Exercise A.5.4 for a determination of G in case of characteristic 0.)

(c) Suppose now that $g \ge 2$. Prove that $\operatorname{Aut}(C)$ is finite. (*Hint.* Show, for example, that an automorphism preserves the set of Weierstrass points.)

A.4.16. (Weil's reciprocity law) Let C be a smooth curve and let $D = \sum n_P P \in \mathrm{Div}(C)$. For $f \in k(C)$ such that the support of D and $\mathrm{div}(f)$ are disjoint, we define

$$f(D) = \prod_{P \in C} f(P)^{n_P}.$$

Let $f, g \in k(C)$ be functions whose divisors with disjoint supports. Prove Weil's reciprocity formula

$$f(\mathrm{div}(g)) = g(\mathrm{div}(f))$$

using the following steps.
(a) Let $\phi : C_1 \to C_2$ be a covering of smooth curves. Prove that if $f \in k(C_1)^*$ and $D \in \mathrm{Div}(C_2)$, then $f(\phi^* D) = (\phi_* f)(D)$. Similarly, prove that if $f \in k(C_2)^*$ and $D \in \mathrm{Div}(C_1)$, then $f(\phi_* D) = (\phi^* f)(D)$.
(b) For the case $C = \mathbb{P}^1$, prove the reciprocity law by a direct computation.
(c) In the general case, use the morphism $g : C \to \mathbb{P}^1$ and part (a) to reduce to the previous case.

A.5. Abelian Varieties over $\mathbb{C}$

We begin by recalling that an abelian variety is an algebraic group that is also a projective variety. With some knowledge of Lie groups, it is easily seen that the set of complex points of such a variety forms a *complex torus* (see Exercise A.5.1). That is, the complex points of an abelian variety are isomorphic to $\mathbb{C}^g/\Lambda$ for some lattice Λ. For example, the complex points of an elliptic curve form a torus $\mathbb{C}/(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2)$. Note, however, that isomorphic means isomorphic as complex-analytic varieties. The isomorphism will be given by holomorphic functions, not by rational functions.

In dimension 1, every complex torus is analytically isomorphic to an abelian variety. It is a somewhat surprising fact that this is not true in dimension greater than or equal to 2. One of the central theorems of this chapter will be the following characterization describing exactly which complex tori $\mathbb{C}^g/\Lambda$ are abelian varieties, that is, which complex tori admit a complex-analytic embedding into some projective space $\mathbb{P}^n(\mathbb{C})$.

Theorem A.5.0.1. *Let Λ be a lattice in $\mathbb{C}^g$. The complex torus $\mathbb{C}^g/\Lambda$ is an abelian variety if and only if there exists a positive definite Hermitian form on $\mathbb{C}^g \times \mathbb{C}^g$ whose imaginary part takes integer values when restricted to $\Lambda \times \Lambda$.*

In light of this statement, we introduce a definition.

Definition. A *Riemann form* with respect to a lattice Λ is a Hermitian form on $\mathbb{C}^g \times \mathbb{C}^g$ whose imaginary part takes integer values when restricted to $\Lambda \times \Lambda$. A Riemann form is called *nondegenerate* if it is positive definite.

Recall that a Hermitian form is a map

$$H : \mathbb{C}^g \times \mathbb{C}^g \to \mathbb{C}$$

that is linear with respect to the first set of variables and satisfies

$$H(z,w) = \overline{H(w,z)}.$$

Theorem A.5.0.1 says that a complex torus can be embedded into a projective space $\mathbb{P}^n$ if and only if it possesses a nondegenerate Riemann form. This chapter is devoted to the proof of Theorem A.5.0.1. Another description of Riemann forms is given by the following easy lemma, whose proof we will leave for the reader.

Lemma A.5.0.2. *Let V be a complex vector space. There is a natural correspondence between Hermitian forms $H : V \times V \to \mathbb{C}$ and real bilinear alternating forms $E : V \times V \to \mathbb{R}$ satisfying $E(ix, iy) = E(x,y)$. This correspondence matches a Hermitian form H with its imaginary part $E = \operatorname{Im}(H)$, and it takes a bilinear alternating form E and attaches it to the Hermitian form $H(x,y) = E(ix,y) + iE(x,y)$.*

Examples A.5.0.3. (a) All tori of dimension one are abelian varieties. Indeed, let $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be a lattice in $\mathbb{C}$. We can define a nondegenerate Riemann form on this lattice by the formula

$$H(z,w) = \frac{z\overline{w}}{\operatorname{Im}(\omega_1\overline{\omega}_2)}.$$

(b) If the dimension of the lattice is greater than one, then there are many tori that do not admit a nonzero Riemann form. For example, let e_1, e_2, e_3, e_4 be vectors in $\mathbb{C}^2$ whose coordinates are all algebraically independent over $\mathbb{Q}$, and let Λ be the lattice that they span. Then the torus $\mathbb{C}^2/\Lambda$ is not an abelian variety.

(c) Let τ be a $g \times g$ symmetric matrix whose imaginary part is positive definite. Then $H(z,w) = {}^t z(\operatorname{Im}\tau)^{-1}\overline{w}$ defines a Riemann form with respect to the lattice $\mathbb{Z}^g + \tau\mathbb{Z}^g$. Hence the torus $\mathbb{C}^g/(\mathbb{Z}^g + \tau\mathbb{Z}^g)$ is an abelian variety. We will use this in the next section to show that the Jacobian variety of a curve is an abelian variety.

(d) The following construction of abelian varieties with complex multiplication (often abbreviated CM) is due to Shimura and Taniyama [1]. Let K_0 be a totally real number field of degree g, let K be a totally imaginary quadratic extension of K_0, and let R_K be the ring of integers of K. The fact that K is totally imaginary means that the embeddings of K in $\mathbb{C}$

come in conjugate pairs. Let $\sigma_1, \ldots, \sigma_g : K \hookrightarrow \mathbb{C}$ be a set of nonconjugate embeddings. We use these embeddings to define a map

$$\Phi : K \longrightarrow \mathbb{C}^g, \qquad x \longmapsto \bigl(\sigma_1(x), \ldots, \sigma_g(x)\bigr).$$

Then $\Lambda = \Phi(R_K)$ is a lattice in $\mathbb{C}^g$. We can construct a nondegenerate Riemann form as follows. Choose an element $\xi \in R_K$ such that $-\xi^2$ is a totally positive element of K_0. Then

$$H(z, w) = 2\sum_{j=1}^{g} \operatorname{Im}\bigl(\sigma_j(\xi)\bigr) z_j \overline{w}_j$$

defines a nondegenerate Riemann form on the torus $\mathbb{C}^g/\Lambda$. To see this, we note that

$$\operatorname{Im} H\bigl(\Phi(x), \Phi(y)\bigr) = \operatorname{Trace}_{\mathbb{Q}}^{K}\bigl(\xi\gamma(x)y\bigr) \in \mathbb{Z},$$

where γ is the nontrivial element of the Galois group G_{K/K_0}.

We also observe that the endomorphism ring $\operatorname{End}(\mathbb{C}^g/\Lambda)$ naturally contains R_K, where the action is induced by multiplication,

$$\alpha(z \bmod \Lambda) = \bigl(\sigma_1(\alpha)z_1, \ldots, \sigma_g(\alpha)z_g\bigr) \bmod \Lambda \qquad \text{for } \alpha \in R_K,\ z \in \mathbb{C}^g.$$

As specific examples we can take $K_0 = \mathbb{Q}\bigl(\sqrt{2}\bigr)$ and $\xi = i$, or $K_0 = \mathbb{Q}\bigl(\cos(2\pi/n)\bigr)$ and $\xi = 2i\sin(2\pi/n)$.

A.5.1. Complex Tori

A *complex torus* T is a compact complex Lie group. In other words, T is a complex manifold of the form V/Λ, where V is a complex vector space and $\Lambda \subset V$ is a lattice of rank $2\dim(V)$. We begin by studying the analytic morphisms between two complex tori T_1 and T_2. By composing an arbitrary morphism with a translation, it suffices to consider morphisms that send the origin of T_1 to the origin of T_2.

Lemma A.5.1.1. *Let $T_1 = V_1/\Lambda_1$ and $T_2 = V_2/\Lambda_2$ be complex tori, and let*

$$\alpha : T_1 \longrightarrow T_2$$

be a holomorphic map with $\alpha(0) = 0$. Then α is a homomorphism that is induced by a $\mathbb{C}$-linear map $\overline{\alpha} : V_1 \to V_2$ satisfying $\overline{\alpha}(\Lambda_1) \subset \Lambda_2$.

PROOF. A complex vector space is simply connected, so the composition $V_1 \to T_1 \xrightarrow{\alpha} T_2$ lifts to a holomorphic map $\overline{\alpha} : V_1 \to V_2$. We necessarily have $\overline{\alpha}(\Lambda_1) \subset \Lambda_2$, and $\overline{\alpha}$ is uniquely determined if we further require that $\overline{\alpha}(0) = 0$.

To see that α is a homomorphism, we write $V_1 = \mathbb{C}^m$ and $V_2 = \mathbb{C}^n$, and we choose coordinate functions $z_1, \ldots, z_m$ for V_1 and $w_1, \ldots, w_n$ for V_2. Then for each $1 \le i \le n$ we can write

$$\overline{\alpha}^*(dw_i) = \sum_{j=1}^{m} a_{ij}(z) dz_j.$$

The fact that $\overline{\alpha}$ is holomorphic means that each of the $a_{ij}(z)$'s is a holomorphic function on $V_1 = \mathbb{C}^m$. On the other hand, the periodicity of $\overline{\alpha}$ means that the a_{ij}'s descend to give functions on the compact space $T_1 = V_1/\Lambda$. The only holomorphic functions on a compact complex manifold are the constant functions (maximum principle), so $a_{ij} \in \mathbb{C}$ for all i, j. It follows that $\overline{\alpha}$ is the linear map defined by the matrix (a_{ij}). □

Remark A.5.1.2. If $T_1 = T_2 = T$, then "multiplication by an integer n" gives an endomorphism of T, which we will denote by $[n]_T$ or $[n]$. In this way we see that $\operatorname{End}(T)$ contains $\mathbb{Z}$. Further, the kernel of multiplication-by-n, which we denote by T_n or $\ker[n]_T$, is given by

$$\ker[n]_T = (1/n)\Lambda/\Lambda \cong (\mathbb{Z}/n\mathbb{Z})^{2\dim T}.$$

So the kernel of multiplication by n is a free $\mathbb{Z}/n\mathbb{Z}$-module of rank $2\dim T$.

Let $\alpha : T_1 \to T_2$ be a holomorphic map as in (A.5.1.1). We observe that the image $\alpha(T_1)$ is again a complex torus, and similarly that the connected component of the kernel of α is a complex torus. It is also immediate that a subtorus of an abelian variety and the homomorphic image of an abelian variety are again abelian varieties.

Example A.5.1.3. Let $T = \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau)$ be a complex torus of dimension one. Then $\operatorname{End}(T) = \mathbb{Z}$ unless τ generates a (necessarily imaginary) quadratic extension of $\mathbb{Q}$. If $[\mathbb{Q}(\tau) : \mathbb{Q}] = 2$, then τ satisfies a relation of the form

$$A\tau^2 + B\tau + C = 0 \qquad \text{with } A, B, C \in \mathbb{Z} \text{ and } \gcd(A, B, C) = 1,$$

and we have $\operatorname{End}(T) = \mathbb{Z} + \mathbb{Z}A\tau$.

More generally, let $T_1 = \mathbb{C}/(\mathbb{Z}+\mathbb{Z}\tau_1)$ and $T_2 = \mathbb{C}/(\mathbb{Z}+\mathbb{Z}\tau_2)$ be complex tori of dimension one. The group $\operatorname{Hom}(T_1, T_2)$ will be nontrivial if and only if there are rational numbers a, b, c, d such that

$$\tau_2 = \frac{a\tau_1 + b}{c\tau_1 + d}.$$

Query A.5.1.4. Let τ be a symmetric $g \times g$ matrix with positive definite imaginary part, and let $T := \mathbb{C}^g/(\mathbb{Z}^g + \tau\mathbb{Z}^g)$. When is it true that $\operatorname{End}(T)$ is strictly larger than $\mathbb{Z}$?

Definition. Let G_1 and G_2 be two algebraic (or analytic) groups. A map $\alpha \in \operatorname{Hom}(G_1, G_2)$ is called an *isogeny* if it is surjective, has finite kernel, and $\dim G_1 = \dim G_2$. The cardinality of $\ker(\alpha)$ is called the *degree* of α. (N.B. This definition of the degree is appropriate only for separable maps.)

It is clear that if G_2 is connected, then two of the defining properties of an isogeny imply the third. Our first examples of isogenies are the multiplication-by-n maps (A.5.1.2). The next lemma says that every isogeny between tori factors through a multiplication map.

Lemma A.5.1.5. *Let T_1 and T_2 be complex tori, and let $\alpha : T_1 \to T_2$ be an isogeny of degree d. There exists a unique isogeny $\hat{\alpha} : T_2 \to T_1$ such that $\alpha \circ \hat{\alpha} = [d]_{T_2}$ and $\hat{\alpha} \circ \alpha = [d]_{T_1}$. The isogeny $\hat{\alpha}$ is called the dual isogeny to α.*

PROOF. By definition, α is surjective, and the definition of its degree implies that $\ker(\alpha) \subset \ker[d]_{T_1}$. It follows that there is a unique $\hat{\alpha}$ such that $\hat{\alpha} \circ \alpha = [d]_{T_1}$. But then

$$\alpha \circ \hat{\alpha} \circ \alpha(x) = \alpha\big([d]_{T_1}(x)\big) = [d]_{T_2}\big(\alpha(x)\big),$$

so we also obtain $\alpha \circ \hat{\alpha} = [d]_{T_2}$. □

Remark A.5.1.6. Let $\alpha : T_1 \to T_2$ be an isogeny of degree d as described in (A.5.1.5), and let $g = \dim(T_1)$. The multiplication map $[d]_{T_1}$ has degree d^{2g}, so we see that $\deg(\hat{\alpha}) = \deg(\alpha)^{2g-1}$. Lemma A.5.1.5 shows that the relation "T_1 is isogenous to T_2" is symmetric. We also observe that the lemma can be proven with d equal to the exponent of $\ker(\alpha)$, rather than its cardinality, albeit with a different $\hat{\alpha}$.

The next theorem is the first that requires the torus to have the structure of an abelian variety. It is not valid for tori in general, since its proof relies on the existence of a nondegenerate Riemann form.

Theorem A.5.1.7. (Poincaré irreducibility theorem) *Let $A = V/\Lambda$ be an abelian variety, and let B be an abelian subvariety of A. Then there exists another abelian subvariety C such that $B + C = A$ and $B \cap C$ is finite. In other words, the map*

$$B \times C \longrightarrow A, \qquad (b, c) \longmapsto b + c,$$

is an isogeny.

PROOF. Let H be a nondegenerate Riemann form for A, and let E be its imaginary part. The tangent space of A is naturally identified with V, and we let $V_1 \subset V$ be the tangent space of B. We also set $\Lambda_1 = V_1 \cap \Lambda$, so $B = V_1/\Lambda_1$. Now consider the orthogonal complement of V_1 with respect to H,

$$V_2 = \big\{ v \in V \mid H(v, w) = 0 \text{ for all } w \in V_1 \big\}.$$

If $w \in V_1$, then also $iw \in V_1$, since V_1 is a complex vector space. By definition we have $H(v,w) = E(v,-iw) + iE(v,w)$, so we can also write V_2 as

$$V_2 = \{v \in V \mid E(v,w) = 0 \text{ for all } w \in V_1\}.$$

Now we look at

$$\Lambda_2 = \Lambda \cap V_2 = \{x \in \Lambda \mid E(x,y) = 0 \text{ for all } y \in \Lambda_1\}.$$

The assumption that E is nondegenerate combined with the fact that Λ_1 is a lattice in V_1 means that Λ_2 has rank

$$\operatorname{rank} \Lambda_2 = \operatorname{rank} \Lambda - \operatorname{rank} \Lambda_1 = 2 \dim_{\mathbb{C}} V_2.$$

Hence Λ_2 is a lattice in V_2, and $C = V_2/\Lambda_2$ is an abelian subvariety of A. Since $V = V_1 \oplus V_2$, we deduce that $B + C = A$ and that $B \cap C$ is finite. □

Definition. A torus is said to be *simple* if it does not contain any nontrivial subtori.

A straightforward consequence of Poincaré's irreducibility theorem is the next result.

Corollary A.5.1.8. *Any abelian variety A is isogenous to a product of the form*

$$A_1^{n_1} \times \cdots \times A_s^{n_s},$$

where the A_i's are simple, pairwise nonisogenous abelian varieties.

PROOF. The proof is by induction on the dimension of A. An abelian variety of dimension 1 is automatically simple. Suppose now that $\dim(A) = d$ and that the theorem has been proven for lower dimensions. If A is simple, we are done. Otherwise, Theorem A.5.1.7 tells us that A is isogenous to a product $B \times C$, and we can apply the induction hypothesis to both factors to conclude the proof. □

Remark A.5.1.9. Let A be a simple abelian variety. Then one can show that its endomorphism ring $\operatorname{End}(A)$ is an order in a division algebra. It follows that $\operatorname{End}(A^n)$ is the ring $\operatorname{Mat}(n, \operatorname{End}(A))$ of $n \times n$ matrices with coefficients in the ring $\operatorname{End}(A)$. Further, nonisogenous simple abelian varieties admit no nontrivial maps from one to another. So if $A = A_1^{n_1} \times \cdots \times A_s^{n_s}$ with the A_i's simple and pairwise nonisogenous as in Corollary A.5.1.8, then $\operatorname{End}(A) = \prod_{i=1}^{s} \operatorname{Mat}(n_i, \operatorname{End}(A_i))$. For more details and a precise description of the possible division algebras, see Mumford [2, Section 19].

A.5.2. Divisors, Theta Functions, and Riemann Forms

The projectivity of a variety is equivalent to the existence of an ample divisor. Therefore, in order to prove Theorem A.5.0.1, we must study divisors on tori. Such divisors are defined similarly to Cartier divisors, but with analytic functions in place of rational functions. Let us denote by p the projection $p : V = \mathbb{C}^g \to V/\Lambda$. Then a divisor on the torus V/Λ induces a divisor p^*D on $V = \mathbb{C}^g$, and this divisor must be invariant under translation by Λ. If one knows Cousin's theorem (which we will not use), one knows that p^*D must be a principal divisor; that is, $p^*D = \operatorname{div}(f)$ for some meromorphic function f. The invariance property of p^*D implies a functional equation of the form $f(z+\lambda) = \exp(g_\lambda(z))f(z)$. Conversely, a function with such a functional equation defines a divisor on V/Λ. Liouville's theorem implies that a Λ-periodic entire function must be constant, so we cannot hope to construct any interesting functions using constant g_λ's. This leads us to take the next simplest sort of functions, which motivates the following definition.

Definition. An entire function f on $\mathbb{C}^g$ is a *theta function* relative to the lattice Λ if it satisfies a functional equation of the form

$$f(z+\lambda) = \exp(g_\lambda(z))f(z) \qquad \text{for all } \lambda \in \Lambda,$$

where g_λ is an affine function of z. That is, $g : \mathbb{C}^g \to \mathbb{C}$ has the property that $g(z+w) + g(0) = g(z) + g(w)$ for all $z, w \in \mathbb{C}^g$.

The function $\exp(g_\lambda(z))$ is sometimes called the *automorphy factor* of the theta function.

Examples A.5.2.1. (a) (Weierstrass sigma function) Let Λ be a lattice in $\mathbb{C}$, let $\Lambda' = \Lambda \smallsetminus \{0\}$, and define

$$\sigma(z) = z \prod_{\lambda\in\Lambda'} \left(1 - \frac{z}{\lambda}\right) \exp\left(\frac{z}{\lambda} + \frac{1}{2}\left(\frac{z}{\lambda}\right)^2\right).$$

It is clear that σ vanishes precisely on Λ with multiplicity 1, so $\operatorname{div}(\sigma) = \Lambda$. Hence σ induces the divisor (0) on the elliptic curve $\mathbb{C}/\Lambda$. Let us verify that σ is a theta function with respect to Λ. Taking the logarithmic derivative yields the Weierstrass zeta function

$$\zeta(z) = \frac{\sigma'(z)}{\sigma(z)} = \frac{1}{z} + \sum_{\lambda\in\Lambda'} \left(\frac{1}{z-\lambda} + \frac{1}{\lambda} + \frac{z}{\lambda^2}\right).$$

Differentiating once more, we obtain the Weierstrass $\wp$-function

$$\wp(z) = -\zeta'(z) = \frac{1}{z^2} + \sum_{\lambda\in\Lambda'} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\right).$$

It is not hard to check that $\wp$ is periodic relative to Λ. Then two integrations of $\wp(z+\lambda)=\wp(z)$ gives first

$$\zeta(z+\lambda)=\zeta(z)+\eta(\lambda), \qquad \text{and then} \qquad \sigma(z+\lambda)=\sigma(z)\exp\bigl(\eta(\lambda)z+a(\lambda)\bigr).$$

Here $\eta(\lambda)$ and $a(\lambda)$ are constants depending on λ that are independent of z. For further details on the Weierstrass σ, ζ, and $\wp$ functions, see, for example, Lang [11], Silverman [1, Chapter VI], or Silverman [2, Chapter I].
(b) (Riemann's theta function) We consider as in Example A.5.0.3(c) a lattice $\Lambda=\mathbb{Z}^g+\tau\mathbb{Z}^g$ in $\mathbb{C}^g$, and we define

$$\theta(z)=\theta(z,\tau)=\sum_{m\in\mathbb{Z}^g}\exp\{\pi i\,{}^t\!m\tau m+2\pi i\,{}^t\!mz\}.$$

Then θ satisfies the functional equation

$$\theta(z+\ell+\tau n)=\theta(z)\exp(-2\pi i\,{}^t\!nz-\pi i\,{}^t\!n\tau n) \qquad \text{for all } z\in\mathbb{C}^g \text{ and } \ell,n\in\mathbb{Z}^g.$$

Indeed, we need merely observe that

$${}^t\!h\tau h+2\,{}^t\!h(z+\ell+\tau n)={}^t\!(h+n)\tau(h+n)+2\,{}^t\!(h+n)z+2\,{}^t\!h\ell-2\,{}^t\!nz-{}^t\!n\tau n$$

and translate the variable of summation in the series by n.

The following fundamental theorem justifies the introduction of these quasi-periodic theta functions. It says that they can be used to represent all divisors on complex tori.

Theorem A.5.2.2. (Poincaré) *Let D be an effective analytic divisor on a complex torus $T=V/\Lambda$. Then there exists an entire theta function with respect to Λ that represents that divisor.*

PROOF. For the general case, we refer the reader to Lang [4, X, Theorem 1.1] or Swinnerton-Dyer [1, II Theorem 18]. We merely note that if $T=\mathbb{C}/\Lambda$ is of dimension one, the proof is easy using the sigma function (A.5.2.1(a)). For in this case a divisor has the form $D=\sum n_iP_i$. So we choose $u_i\in\mathbb{C}$ such that $P_i=u_i \bmod \Lambda$, and then the function $\theta(z):=\prod\sigma(z-u_i)^{n_i}$ is a theta function that induces the divisor D. □

One might ask to what extent a divisor D determines a theta function. The next lemma answers this question.

Lemma A.5.2.3. *Let θ_1 and θ_2 be theta functions with respect to a lattice Λ, and suppose that they define the same divisor. Then there exists a quadratic form Q, a linear form R, and a constant S such that $\theta_1(z)/\theta_2(z)=\exp\bigl(Q(z)+R(z)+S\bigr)$.*

PROOF. The function $\theta_1(z)/\theta_2(z)$ is an entire nonvanishing function, so we can write it as $\exp\bigl(f(z)\bigr)$ for some entire function f. Applying the functional equations of θ_1 and θ_2, we find that

$$\exp\bigl(f(z+\lambda)-f(z)\bigr)=\exp\bigl(L_\lambda(z)\bigr),$$

where $L_\lambda(z)$ is an affine function of z. It follows that $f(z+\lambda) - f(z)$ is also affine in z. Therefore, all second-order derivatives of f are Λ-periodic and entire, hence constant. This proves that f is a polynomial of degree at most 2, so it can be written in the form $f(z) = Q(z) + R(z) + S$, with Q a quadratic form, R a linear form, and S a constant. □

Definition. A theta function of the form $\exp\big(Q(z) + R(z) + S\big)$, where Q is a quadratic form, R is a linear form, and S is a constant, will be called a *trivial theta function.*

We will now reverse the above procedure and associate to each theta function a Riemann form. To ease notation, we set

$$\mathbf{e}(z) = \exp(2\pi i z).$$

The functional equation of a theta function θ with respect to the lattice Λ can be written as

$$\theta(z+\lambda) = \theta(z)\,\mathbf{e}\big(L(z,\lambda) + J(\lambda)\big),$$

where $L(z,\lambda)$ is a linear function of z. We use this formula to expand $\theta(z+\lambda+\mu)/\theta(z)$ in two ways, which yields the result

$$L(z,\lambda+\mu) - L(z,\mu) - L(z,\lambda) - L(\lambda,\mu) + J(\lambda+\mu) - J(\lambda) - J(\mu) \in \mathbb{Z} \quad \text{for all } (z,\lambda,\mu) \in \mathbb{C}^g \times \Lambda \times \Lambda.$$

By continuity we further deduce that:

(1) $L(z,\lambda+\mu) = L(z,\mu) + L(z,\lambda)$.
(2) $J(\lambda+\mu) - J(\lambda) - J(\mu) \equiv L(\lambda,\mu) \pmod{\mathbb{Z}}$.
(3) $L(\lambda,\mu) \equiv L(\mu,\lambda) \pmod{\mathbb{Z}}$.

From (1) we see that $L(z,\lambda)$ is $\mathbb{Z}$-linear in λ, so it can be extended $\mathbb{R}$-linearly in the second variable to give a map

$$L : V \times V \longrightarrow \mathbb{C}.$$

In other words, $L(z,w)$ is $\mathbb{C}$-linear in z and $\mathbb{R}$-linear in w. From (3) we deduce that the form $E(z,w) = L(z,w) - L(w,z)$ takes integral values on $\Lambda \times \Lambda$. But E is $\mathbb{R}$-linear and $V = \Lambda \otimes \mathbb{R}$, so we see that E is, in fact, real-valued.

Next we observe that

$$\begin{aligned} E(iz,iw) - E(z,w) &= L(iz,iw) - L(iw,iz) - L(z,w) + L(w,z) \\ &= iL(z,iw) - iL(w,iz) + iL(iz,w) - iL(iw,z) \\ &= i\big(E(z,iw) + E(iz,w)\big). \end{aligned}$$

But E takes on only real values, so we must have $E(iz,iw) = E(z,w)$. We summarize this discussion in the next proposition.

Proposition A.5.2.4. *Let θ be a theta function with respect to the lattice Λ, and write the functional equation of θ as*

$$\theta(z+\lambda) = \theta(z)\,\mathbf{e}\big(L(z,\lambda)+J(\lambda)\big).$$

Then the formulas

$$E(z,w) = L(z,w) - L(w,z) \qquad \text{and} \qquad H(z,w) = E(iz,w) + iE(z,w)$$

define a Riemann form with respect to the lattice Λ. Further, H depends only on the divisor of θ, and if we denote by H_D the Riemann form associated to a divisor D, then we have the addition law $H_{D+D'} = H_D + H_{D'}$.

PROOF. Only the last assertion remains to be proven, and it follows from a straightforward computation. □

The most important property of the Riemann form associated to a theta function is given by the next proposition.

Proposition A.5.2.5. (a) *The Riemann form H associated to a theta function is positive.*
(b) *Let $W = \ker H \subset V$. Then the form H' induced by H on V/W is positive definite. Let Λ' be the image of Λ in V/W. Then the function θ is constant on cosets $w+W$, and it induces a theta function with respect to the lattice Λ' whose associated Riemann form is H'.*

PROOF. We start by multiplying the given theta function by a trivial one in order to obtain a nicer functional equation.

Lemma A.5.2.6. *Let θ_0 be a theta function with respect to a lattice Λ, and let H be its Riemann form. Then there exists a theta function θ with the same divisor (and Riemann form) such that*

$$\theta(z+\lambda) = \exp\left(\pi H(z,\lambda) + \frac{\pi}{2}H(\lambda,\lambda) + 2\pi i K(\lambda)\right)\theta(z),$$

where $K : \Lambda \to \mathbb{R}$ is a function satisfying the identity

$$\mathbf{e}\big(K(\lambda+\mu)\big) = \mathbf{e}\big(K(\lambda)\big)\,\mathbf{e}\big(K(\mu)\big)\,\mathbf{e}\left(\frac{1}{2}E(\lambda,\mu)\right).$$

Furthermore, there is a constant $C = C(\theta)$ such that the function θ satisfies the growth estimate

$$\big|\theta(z)\big| \le C\exp\left(\frac{\pi}{2}H(z,z)\right).$$

PROOF. Let Q be a bilinear form. We are going to consider the function $\theta_1(z) = \exp\bigl(Q(z,z)\bigr)\theta_0(z)$. If the function $\mathbf{e}\bigl(L_0(z,\lambda) + J_0(\lambda)\bigr)$ is the automorphy factor of θ_0, then the automorphy factor of θ_1 will have the form

$$\mathbf{e}\bigl(L(z,\lambda) + J(\lambda)\bigr) \qquad \text{with} \qquad L(z,\lambda) = L_0(z,\lambda) + 2Q(z,\lambda).$$

Notice that by varying the bilinear form Q, we are able to obtain all linear forms L that satisfy $L(z,w) - L(w,z) = E(z,w)$. Indeed, this condition is necessary, and if it is satisfied, then $L - L_0$ is symmetric. But since H is Hermitian and $E = \operatorname{Im} H$, we know that

$$\frac{1}{2i}H(z,w) - \frac{1}{2i}\overline{H(z,w)} = E(z,w).$$

Therefore, we may select Q such that $L(z,w) = \frac{1}{2i}H(z,w)$. This means that if we set $K_1(\lambda) = J(\lambda) - \frac{1}{4i}H(\lambda,\lambda)$, then the functional equation of θ_1 can be rewritten as

$$\theta_1(z+\lambda) = \exp\left(\pi H(z,\lambda) + \frac{1}{2}\pi H(\lambda,\lambda) + 2\pi i K_1(\lambda)\right)\theta_1(z).$$

Multiplying θ_1 by $\mathbf{e}\bigl(R(z)\bigr)$ with R linear will not change L and will replace $K_1(\lambda)$ by $K(\lambda) = K_1(\lambda) + R(\lambda)$. Now using relation (2) between J and $L = \frac{1}{2i}H$, we find that

$$K_1(\lambda+\mu) - K_1(\lambda) - K_1(\mu) \equiv \frac{1}{2}E(\lambda,\mu) \pmod{\mathbb{Z}}.$$

Hence we may assume that $\operatorname{Im} K_1$ is $\mathbb{Z}$-linear and extend it to V by $\mathbb{R}$-linearity. Taking $R(z) = -\operatorname{Im} K_1(iz) - \operatorname{Im} K_1(z)$, we obtain a $\mathbb{C}$-linear function such that $K(\lambda) = K_1(\lambda) + R(\lambda)$ is real. This gives the first part of the lemma. To prove the second part, we merely need to observe that the function

$$\bigl|\theta(z)\bigr| \exp\left(-\frac{\pi}{2}H(z,z)\right)$$

is Λ-periodic and continuous, hence bounded. □

We return to the proof of Proposition A.5.2.5. Suppose that there exists some $v_0 \in V$ with $H(v_0,v_0) < 0$. For every $z \in \mathbb{C}$ we have the estimate

$$\bigl|\theta(zv_0)\bigr| \le C\exp\left(\frac{\pi}{2}|z|^2 H(v_0,v_0)\right)$$

from (A.5.2.6), so we see that $\theta(zv_0) \to 0$ as $z \to \infty$. From Liouville's theorem we conclude that $\theta(zv_0) = 0$. But by continuity, the inequality $H(v,v) < 0$ must remain true for all v in a small neighborhood of v_0. This

would imply that θ vanishes identically. Hence the Riemann form H is positive, which proves (a).

For (b), we note that if $w \in W$, then

$$|\theta(z+w)| \leq C \exp\left(\frac{\pi}{2}H(z+w,z+w)\right) = C \exp\left(\frac{\pi}{2}H(z,z)\right).$$

Hence applying Liouville again, we see that $\theta(z+w) = \theta(z)$, and the rest of the proposition is just linear algebra. □

Let θ be a theta function with divisor D for a lattice Λ in $V = \mathbb{C}^g$. We write $L(\theta)$ for the vector space of all theta functions with the same functional equation. We will see in the next section that $L(\theta)$ is finite-dimensional, so choosing a basis $\theta_0, \dots, \theta_n$ for $L(\theta)$, we get a holomorphic map

$$\phi_D : V/\Lambda \longrightarrow \mathbb{P}^n(\mathbb{C}), \qquad z \bmod \Lambda \longmapsto \big(\theta_0(z), \dots, \theta_n(z)\big).$$

In order to state the next theorem, which clearly implies Theorem A.5.0.1, we introduce the following ad hoc definition.

Definition. The divisor D on the torus V/Λ is *very ample* if the map ϕ_D described above is an embedding. The divisor D is *ample* if some positive multiple of D is very ample.

Notice that if we know, a priori, that V/Λ is an algebraic variety, then this definition of ample coincides with the one given in Section A.3. To ease notation, we may take the given theta function θ to be θ_0. Then for every $f \in L(D)$, the function $\theta(z)f(z)$ is entire and belongs to $L(\theta)$. Conversely, each $f_i = \theta_i/\theta$ is an abelian function with $D + (f_i) \geq 0$. Hence $L(D)$ and $L(\theta)$ are isomorphic, and we may speak of the linear system associated to a theta function.

Theorem A.5.2.7. *Let D be an effective divisor on a torus. The Riemann form attached to D is nondegenerate if and only if D is ample.*

PROOF. If the Riemann form H attached to D is degenerate, then by Proposition A.5.2.5, all of the theta functions in $L(\theta)$ are constant on the cosets $w + \ker H$. It follows that the map $\phi_D : V/\Lambda \to \mathbb{P}^n$ cannot be an embedding.

Suppose now that H is nondegenerate. To prove that ϕ_{mD} is an embedding, we must construct theta functions $\theta_0, \dots, \theta_n$ associated to mH such that $\theta_0, \dots, \theta_n$ have no common zeros and separate tangent vectors. This will be done in the next section using a theorem of Lefschetz and a Riemann–Roch theorem for complex tori. □

A.5.3. Riemann–Roch for Abelian Varieties

The purpose of this section is to compute the dimension of the linear system $L(\theta)$ and to finish the proof that a divisor with a positive definite Riemann form is ample. The computation of $\ell(D)$ involves the determinant of the associated alternating form. We recall how this quantity is computed.

Lemma A.5.3.1. (Frobenius) *Let Λ be a free abelian group of rank $2g$ (i.e., $\Lambda \cong \mathbb{Z}^{2g}$). Let E be a nondegenerate bilinear alternating form on Λ with values in $\mathbb{Z}$. There exist positive integers $d_1, \dots, d_g$ with $d_i | d_{i+1}$ and a basis $\mathbf{e}_1, \dots, \mathbf{e}_g, \mathbf{f}_1, \dots, \mathbf{f}_g$ of Λ such that*

$$E(\mathbf{e}_i, \mathbf{e}_j) = E(\mathbf{f}_i, \mathbf{f}_j) = 0 \qquad \text{and} \qquad E(\mathbf{e}_i, \mathbf{f}_j) = \begin{cases} d_i & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The product $d_1 \cdots d_g$ is the square root of the determinant of E.

PROOF. We use induction on g. Note that the set $\{E(x, y) \mid x, y \in \Lambda\}$ is an ideal of $\mathbb{Z}$, so it is a principal ideal, generated by some positive integer $d_1 = E(\mathbf{e}_1, \mathbf{f}_1)$. Applying the induction hypothesis to the orthogonal complement of $\mathbb{Z}\mathbf{e}_1 + \mathbb{Z}\mathbf{f}_1$ in Λ finishes the proof. □

Definition. Let E be a nondegenerate bilinear alternating form on $\Lambda \cong \mathbb{Z}^{2g}$ with values in $\mathbb{Z}$. A basis $\mathbf{e}_1, \dots, \mathbf{e}_g, \mathbf{f}_1, \dots, \mathbf{f}_g$ for Λ as in Lemma A.5.3.1 is called a *Frobenius basis*, and the d_i's are called the *invariants* of E. Further, we define the *Pfaffian of E* to be the quantity

$$\mathrm{Pf}(E) = d_1 d_2 \cdots d_g = \sqrt{\det(E)}.$$

Our next task is to modify a given theta function to produce the simplest possible functional equation.

Lemma A.5.3.2. *Let θ be a theta function with nondegenerate Riemann form H on the lattice $\Lambda \subset V = \mathbb{C}^g$. Let $\{\mathbf{e}_1, \dots, \mathbf{e}_g, \mathbf{f}_1, \dots, \mathbf{f}_g\}$ be a Frobenius basis for the form $E = \operatorname{Im} H$ on Λ, and let $d_1, \dots, d_g$ be the associated invariants.*
(a) *The set $\{\mathbf{e}_1, \dots, \mathbf{e}_g\}$ is a $\mathbb{C}$-basis of V.*
(b) *After multiplication by a suitable trivial theta function, the functional equation of θ takes the form*

$$\theta(z + \mathbf{e}_i) = \theta(z) \qquad \text{and} \qquad \theta(z + \mathbf{f}_i) = \theta(z)\, \mathbf{e}(d_i z_i + c_i).$$

We use (a) *to write $z = \sum z_i \mathbf{e}_i$.*

PROOF. (a) Let $W = \mathbb{R}\mathbf{e}_1 + \cdots + \mathbb{R}\mathbf{e}_g \subset V$. Then E vanishes on W, so if $x, y \in W$ and $x + iy = 0$, then $iy \in W$ and $E(iy, y) = 0$. This implies that $y = x = 0$, and hence that $V = W + iW$.

(b) The functional equation reads $\theta(z+\lambda)=\theta(z)\,\mathbf{e}\big(L(z,\lambda)+J(\lambda)\big)$. But for $x,y\in W$, we know that $L(x,y)-L(y,x)=E(x,y)=0$, so L is symmetric on $W\times W$. We define a bilinear form Q by requiring that $Q(\mathbf{e}_i,\mathbf{e}_j)=-\frac{1}{2}L(\mathbf{e}_i,\mathbf{e}_j)$, and we multiply θ by $\mathbf{e}\big(Q(z,z)\big)$. The new theta function will have an L satisfying

$$L(\mathbf{e}_i,\mathbf{e}_j)=0 \qquad \text{and} \qquad L(\mathbf{e}_i,\mathbf{f}_j)=L(\mathbf{f}_j,\mathbf{e}_i)+E(\mathbf{e}_i,\mathbf{f}_j)=\begin{cases} d_i & \text{if } i=j,\\ 0 & \text{if } i\neq j.\end{cases}$$

In other words, $L(z,\mathbf{e}_i)=0$, and if we write $z=\sum z_j\mathbf{e}_j$, then $L(z,\mathbf{f}_i)=d_iz_i$. Finally, we know that

$$J(\lambda+\mathbf{e}_i)-J(\lambda)-J(\mathbf{e}_i)=L(\lambda,\mathbf{e}_i)=0 \pmod{\mathbb{Z}}.$$

So if we let R be the linear form determined by $R(\mathbf{e}_i)=-J(\mathbf{e}_i)$ and multiply θ by $\mathbf{e}\big(R(z)\big)$, we obtain a new function J such that $J(\mathbf{e}_i)=0$. The lemma is then proven with $c_i=J(\mathbf{f}_i)$. □

Theorem A.5.3.3. (Riemann–Roch for abelian varieties) *Let D be a divisor on an abelian variety, let H_D be the nondegenerate Riemann form for D, let $E_D=\operatorname{Im}H_D$, and let $\operatorname{Pf}(E_D)$ be its Pfaffian. Then $\ell(D)=\operatorname{Pf}(E_D)$.*

PROOF. The proof is a simple consequence of the results proven above. We may suppose that D is defined by a theta function θ with functional equation as in Lemma A.5.3.2. Note that any theta function with the same functional equation as θ is will be periodic with respect to $\Lambda=\mathbb{Z}\mathbf{e}_1+\cdots+\mathbb{Z}\mathbf{e}_g$. Identifying Λ with $\mathbb{Z}^g$, we can expand θ as a Fourier series,

$$\theta(z)=\sum_{m\in\mathbb{Z}^g} a(m)\,\mathbf{e}({}^t\!mz).$$

The second half of the functional equation gives a recursion formula for the coefficients $a(m)$, namely

$$a(m-d_i\mathbf{e}_i)=a(m)\,\mathbf{e}({}^t\!m\mathbf{f}_i-c_i).$$

It is immediate that all of the $a(m)$'s are determined uniquely by the values of $a(m)$ for

$$m\in\big\{(m_1,\ldots,m_g)\in\mathbb{Z}^g \mid 0\le m_i\le d_i-1\big\}.$$

This set has cardinality $d_1\cdots d_g$, which shows that $\ell(D)\le\operatorname{Pf}(E_D)$. In order to show that we have equality, it remains only to show that each choice of $a(m)$'s with m in this set leads, via the above recursion, to a Fourier series that converges. This fact is an easy consequence of the following lemma.

Lemma A.5.3.4. *With notation as above, set*

$$Q(n,n) = -\frac{1}{2}L\left(\sum n_i \mathbf{f}_i, \sum n_i \mathbf{f}_i\right).$$

Then the coefficients $a(m)$ *have the form*

$$a\left(n_0 + \sum n_i d_i \mathbf{e}_i\right) = a(n_0)\,\mathbf{e}\big(Q(n,n) + R(n)\big),$$

where R *is linear and the imaginary part of* Q *is positive definite.*

PROOF. We leave the first part as an exercise and just prove that $\operatorname{Im} Q > 0$. Let $z \in \mathbb{R}\mathbf{f}_1 + \cdots + \mathbb{R}\mathbf{f}_g$. We write $z = x + iy$ with $x, y \in \mathbb{R}\mathbf{e}_1 + \cdots + \mathbb{R}\mathbf{e}_g$. Then $L(z,z) = L(x,z) + iL(y,z)$, where $L(x,z) = E(x,z)$ and $L(y,z) = E(y,z)$ are real. Hence we obtain

$$\operatorname{Im}\big(-L(z,z)\big) = -L(y,z) = E(z,y) = E(x,y) + E(iy,y) = H(y,y) > 0,$$

as was to be shown. □

We give two consequences that will be used in the proof of Theorem A.5.3.6 below.

Corollary A.5.3.5. *Let* θ_0 *be a theta function for the lattice* Λ, *and assume that the form associated to* θ_0 *is nondegenerate.*
(i) *There exists a theta function in* $L(\theta_0)$ *that is not a theta function for any lattice strictly containing* Λ.
(ii) *Every* $\theta \in L(\theta_0)$ *depends on each of the variables* $z_1, \dots, z_g$.

PROOF. Suppose that $\Lambda' \supset \Lambda$ with $\Lambda' \neq \Lambda$. The set of $\theta \in L(\theta_0)$ that are theta functions for Λ' form a subspace whose dimension is some Pfaffian which strictly divides the Pfaffian of θ_0. But there are only countably many lattices containing Λ, and a countable union of proper subspaces cannot fill a (complex) vector space. This proves (i).

For (ii), we suppose that $\partial\theta/\partial z_1 = 0$. Then $L(z,\lambda)$ does not depend on z_1; hence $L(\mathbf{e}_1, w) = 0$ and $E(i\mathbf{e}_1, \mathbf{e}_1) = H(\mathbf{e}_1, \mathbf{e}_1) = 0$. This contradicts the nondegeneracy. □

Finally, we prove the fundamental embedding theorem of Lefschetz.

Theorem A.5.3.6. (Lefschetz) *Let* θ *be a theta function with divisor* D *and nondegenerate Riemann form* H.
(i) *The divisor* $2D$ *is base point free.*
(ii) *The divisor* $3D$ *is very ample.*

PROOF. Note that this gives a broad generalization of the corresponding result on elliptic curves. The proof of (i) is easy and only requires D to

be effective. To ease notation, we write $\theta_w(z) = \theta(z-w)$. Notice that $\theta_w(z)\theta_{-w}(z) \in L(\theta^2)$. If z_0 were a base point of θ^2, then we would have

$$\theta_w(z_0)\theta_{-w}(z_0) = 0 \qquad \text{for all } w.$$

But this is impossible unless $\theta = 0$. Hence $2D$ is base point free.

Next we observe that for any u, v, the product $\theta_u\theta_v\theta_{-u-v}$ is in $L(\theta^3)$. We are going to prove that these theta functions span a very ample linear system. Let $\Phi : \mathbb{C}^g/\Lambda \to \mathbb{P}L(\theta^3)$ be the holomorphic map associated to $L(\theta^3)$. Using Corollary A.5.3.5, we may assume that θ is not quasi-periodic with respect to any lattice strictly larger than Λ.

We begin by showing that Φ is injective. Suppose that $\Phi(x) = \Phi(y)$. Then for every u, v there is some $a \in \mathbb{C}^*$ such that

$$(\theta_u\theta_v\theta_{-u-v})(x) = a(\theta_u\theta_v\theta_{-u-v})(y).$$

This implies that $\theta(x-u) = \theta(y-u)g(u)$ for some nonvanishing entire function g. But g must then be a trivial theta function and must satisfy $g(u+\lambda) = \mathbf{e}\big(L(x-y,\lambda)\big)g(u)$ for all $\lambda \in \Lambda$. One easily gets from this that $g(u) = g(0)\,\mathbf{e}\big(-L(u,x-y)\big)$, and this in turn implies $\theta\big(z+(x-y)\big) = \theta(z)\,\mathbf{e}\big(-L(z,x-y)+c\big)$. Hence θ is a theta function with respect to the lattice $\Lambda + \mathbb{Z}(x-y)$. However, we know that Λ is the largest lattice for which θ is a theta function, so we have proven that $x - y \in \Lambda$. This completes the proof that Φ is injective.

In order to prove that Φ is an embedding, it remains to show that it separates tangent directions. Suppose to the contrary that the differential of Φ annihilates a tangent vector at some point w. We may assume that $\theta(w) \neq 0$, and after a change of coordinates we may assume that all of the functions $f = \theta_u\theta_v\theta_{-u-v}/\theta^3$ satisfy $\partial f/\partial z_1(w) = 0$. Let $r = \theta^{-1}(\partial\theta/\partial z_1)$. Taking the logarithmic derivative of f at w, we find that

$$r(w-u) + r(w-v) + r(w+u+v) - 3r(w) = 0.$$

This implies that $r(u+w) = a_0 + a_1u_1 + \cdots + a_gu_g$ is an affine function of u, and then integrating with respect to u_1 gives a new theta function θ' such that

$$\theta(u+w)\exp(-(a_1/2)u_1^2) = \theta'(u_2,\ldots,u_g).$$

But the existence of this theta function contradicts Corollary A.5.3.5(ii). This completes the proof that Φ is an embedding. □

EXERCISES

A.5.1. Let $G \subset \mathbb{P}^n(\mathbb{C})$ be a complex projective variety with a group law given by algebraic functions (i.e., G is a projective group variety). This exercise sketches a proof that G is analytically isomorphic to a complex torus. In particular, the compactness of G implies that the group law is abelian.
(a) For each $g \in G$, consider the conjugation map $\phi(g) : G \to G$, $\phi(g)(h) = ghg^{-1}$. Denote the differential of $\phi(g)$ at the identity e by $D(g) : T_e(G) \to T_e(G)$. Using the maximum principle (from complex function theory), show that $D(g)$ is the identity map. Use this to deduce that $\phi(g)$ is also the identity, and hence that G is abelian.
(b) Prove that the exponential map $T_e(G) \to G$ is surjective with discrete kernel Λ. Use the fact that G is compact to conclude that $\operatorname{rank}\Lambda = 2\dim G$, and hence that G is isomorphic to a complex torus. (For basic properties of the exponential map, see, for example, Bourbaki [1, Chapter III].)

A.5.2. Let Ω be a $g \times 2g$ matrix with coefficients in $\mathbb{C}$. Prove that the existence of a nondegenerate Riemann form on the torus $A = \mathbb{C}^g/\Omega\mathbb{Z}^{2g}$ is equivalent to the existence of a nondegenerate alternating $2g \times 2g$ matrix J with coefficients in $\mathbb{Z}$ satisfying

$$\Omega J^{-1}\,{}^t\Omega = 0 \qquad \text{and} \qquad i\Omega J^{-1}\,{}^t\bar{\Omega} > 0.$$

Here $S > 0$ means that the matrix S is symmetric and positive definite.

A.5.3. Use the previous exercise to construct a torus of dimension 2 that is not an abelian surface. (*Hint.* Show that the existence of a Riemann form implies that that coefficients of the matrix of periods satisfy nontrivial relations over $\mathbb{Z}$.)

A.5.4. Let $E = \mathbb{C}/\Lambda$ be a complex torus of dimension 1 (i.e., a complex elliptic curve) and let

$$\operatorname{Aut}(E,0) = \{\text{analytic isomorphisms } \phi : E \to E \text{ with } \phi(0) = 0\}.$$

In other words, G is the group of analytic automorphisms of the group variety E. Also, let $\rho = e^{2\pi i/3}$ be a primitive cube root of unity. Prove that

$$\operatorname{Aut}(E,0) = \begin{cases} \mathbb{Z}/6\mathbb{Z} & \text{if } E \cong \mathbb{C}/(\mathbb{Z}+\mathbb{Z}\rho), \\ \mathbb{Z}/4\mathbb{Z} & \text{if } E \cong \mathbb{C}/(\mathbb{Z}+\mathbb{Z}i), \\ \mathbb{Z}/2\mathbb{Z} & \text{otherwise.} \end{cases}$$

In particular, $\operatorname{Aut}(E,0)$ is always finite. Do there exist abelian varieties of dimension greater than 1 with infinite automorphism groups?

A.5.5. (Theorem of Appell-Humbert and the dual abelian variety) Let $A = V/\Lambda$ be a complex torus, and denote by $\operatorname{NS}(V/\Lambda)$ the group of Riemann forms on V/Λ. Also, let $\mathrm{S}^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ be the circle group. We say that a map $\chi : \Lambda \to \mathrm{S}^1$ is a *semicharacter* for the Riemann form H if it satisfies the functional equation

$$\chi(\lambda+\mu) = \chi(\lambda)\chi(\mu)\,\mathbf{e}\left(\frac{1}{2}\operatorname{Im} H(\lambda,\mu)\right) \qquad \text{for all } \lambda,\mu \in \Lambda.$$

We also define a group

$$P(V/\Lambda) = \{(H,\chi) \mid H \in \mathrm{NS}(V/\Lambda) \text{ and } \chi \text{ is a semicharacter for } H\}.$$

The group law on $P(V/\Lambda)$ is $(H_1,\chi_1)\cdot(H_2,\chi_2) = (H_1+H_2,\chi_1\chi_2)$.
(a) Show that Theorem A.5.2.2 and Lemma A.5.2.6 associate to each divisor D an element $(H_D,\chi_D) \in P(V/\Lambda)$. Check that this induces an isomorphism from $\mathrm{Pic}(V/\Lambda)$ to $P(V/\Lambda)$.
(b) Show that under this isomorphism, the subgroup $\mathrm{Pic}^0(V/\Lambda)$ contained in $\mathrm{Pic}(V/\Lambda)$ is naturally identified with $\mathrm{Hom}(\Lambda, \mathrm{S}^1)$.
(c) Let $\hat{V} = \mathrm{Hom}_{\bar{\mathbb{C}}}(V,\mathbb{C})$ be the space of $\mathbb{C}$-antilinear forms on V (i.e., $\ell(zv) = \bar{z}\ell(v)$). For $\ell \in \hat{V}$ and $v \in V$, let $\langle \ell, v\rangle = \mathrm{Im}(\ell(v))$, and define $\hat{\Lambda} = \{\ell \in \hat{V} \mid \langle \ell, \Lambda\rangle \in \mathbb{Z}\}$. Prove that the map

$$\hat{V} \longrightarrow \mathrm{Hom}(\Lambda,\mathrm{S}^1), \qquad \ell \longmapsto \mathbf{e}(\langle \ell, \cdot\rangle),$$

is a surjective homomorphism with kernel $\hat{\Lambda}$. Further, prove that $\hat{\Lambda}$ is a lattice in $\hat{V}$, and use the resulting isomorphism $\hat{V}/\hat{\Lambda} \to \mathrm{Hom}(\Lambda,\mathrm{S}^1)$ to deduce that $\mathrm{Hom}(\Lambda,\mathrm{S}^1)$ has the structure of a complex torus.
(d) Show that if $A = V/\Lambda$ is a complex abelian variety, then the torus $\hat{A} = \hat{V}/\hat{\Lambda}$ described in (c) is also an abelian variety. It is called the *dual abelian variety* and is isomorphic to $\mathrm{Pic}^0(A)$.
(e) Let A be an abelian variety, and let $D \in \mathrm{Pic}(A)$. For each $a \in A$, let $t_a : A \to A$ denote the translation map $t_a(x) = x + a$. Define a map

$$\Phi_D : A \longrightarrow \mathrm{Pic}(A), \qquad \Phi_D(a) = \mathrm{Cl}(t_a^* D - D).$$

Prove that Φ_D has the following properties: (i) The image of Φ_D lies in $\mathrm{Pic}^0(A) = \hat{A}$. (ii) The map Φ_D depends only on $D \bmod \mathrm{Pic}^0(A)$. (iii) If the associated Riemann form H_D is positive definite (i.e., if D is ample), then Φ_D is an isogeny from A to $\hat{A}$.

A.5.6. (Poincaré divisor) In the previous exercise we associated to a complex abelian variety $A = V/\Lambda$ its dual abelian variety $\hat{A} = \hat{V}/\hat{\Lambda}$. We now define a pairing on $V \times \hat{V}$ by

$$H : (V\times\hat{V})\times(V\times\hat{V}) \to \mathbb{C}, \qquad H\big((v_1,\ell_1),(v_2,\ell_2)\big) = \overline{\ell_2(v_1)} + \ell_1(v_2).$$

(a) Prove that H is a Riemann form with respect to $\Lambda\times\hat{\Lambda}$ and that $\chi(\lambda,\ell) = \mathbf{e}\left(\frac{1}{2}\mathrm{Im}(\ell(\lambda))\right)$ is a semicharacter for H. It follows from Exercise A.5.5 above that the pair (H,χ) defines a divisor class $\mathcal{P} \in \mathrm{Pic}(A\times\hat{A})$. The divisor class $\mathcal{P}$ is called the *Poincaré divisor class.*
(b) Let $i_{\hat{a}} : A \to A\times\hat{A}$ be the map $i_{\hat{a}}(a) = (a,\hat{a})$. Prove that $i_{\hat{a}}^*\mathcal{P}$ is the divisor class on A corresponding to $\hat{a}$.
(c) Let $i_0 : \hat{A} \to A\times\hat{A}$ be the map $i_0(\hat{a}) = (0,\hat{a})$. Prove that $i_0^*\mathcal{P} = 0$.
(d) Prove that the Poincaré divisor class $\mathcal{P}$ is uniquely characterized by the two properties described in (b) and (c). (For further properties of the Poincaré divisor class, see Section A.7.)

A.5.7. Let $A = V/\Lambda$ and $B = V/\Lambda'$ be two complex abelian varieties of dimensions g and h, respectively. By Lemma A.5.1.1, any homomorphism $\alpha : A \to B$ lifts to a $\mathbb{C}$-linear map $\tilde{\alpha} : V \to V'$ satisfying $\tilde{\alpha}(\Lambda) \subset \Lambda'$. We thus obtain two representations

$$\rho_{\mathbb{C}} : \operatorname{Hom}(A, B) \to \operatorname{Hom}_{\mathbb{C}}(V, V') \quad \text{and} \quad \rho_{\mathbb{Q}} : \operatorname{Hom}(A, B) \to \operatorname{Hom}_{\mathbb{Z}}(\Lambda, \Lambda'),$$

called respectively the *complex representation* and the *rational representation* of $\operatorname{Hom}(A, B)$.

(a) Show that the two representations $\rho_{\mathbb{C}}$ and $\rho_{\mathbb{Q}}$ are related by

$$\rho_{\mathbb{Q}} \otimes \mathbb{C} \cong \rho_{\mathbb{C}} \oplus \overline{\rho_{\mathbb{C}}},$$

where $\bar{\rho}$ denotes the complex conjugate of ρ.

(b) For every integer m, prove that α maps A_m into B_m, where recall that A_m denotes the kernel of the multiplication-by-m map on A, and similarly for B_m. Take the inverse limit of the maps

$$\alpha : A_{p^n} \longrightarrow B_{p^n} \quad \text{as } n \to \infty$$

to obtain a homomorphism $\mathbb{Z}_p^{2g} \to \mathbb{Z}_p^{2h}$. (*Hint.* Recall that $A_m \cong (\mathbb{Z}/m\mathbb{Z})^{2g}$ and $B_m \cong (\mathbb{Z}/m\mathbb{Z})^{2h}$, and use the definition of the p-adic integers). Tensoring with $\mathbb{Q}_p$, we obtain a homomorphism $\rho_{\mathbb{Q}_p} : \mathbb{Q}_p^{2g} \to \mathbb{Q}_p^{2h}$. The representation

$$\rho_{\mathbb{Q}_p} : \operatorname{Hom}(A, B) \longrightarrow \operatorname{Hom}_{\mathbb{Q}_p}(\mathbb{Q}_p^{2g}, \mathbb{Q}_p^{2h})$$

is called the *p-adic representation* of $\operatorname{Hom}(A, B)$. Prove that

$$\rho_{\mathbb{Q}} \otimes \mathbb{Q}_p \cong \rho_{\mathbb{Q}_p}.$$

(The importance of the p-adic representation is that it exists over any base field of characteristic $\neq p$.)

(c) Let $A = B$ and let $\alpha : A \to A$ be an isogeny. Prove that $\deg(\alpha) = \det(\rho_{\mathbb{Q}}(\alpha))$.

A.5.8. Let A be a simple abelian variety (i.e., an abelian variety containing no abelian subvarieties other than $\{0\}$ and itself). Let D be a nonzero effective divisor on A. Prove that D is ample. Show that the conclusion may be false if A is not assumed to be simple.

A.5.9. Let $\theta(z)$ be Riemann's theta function as described in Example A.5.2.1. Show that the associated normalized theta function (in the sense of Lemma A.5.2.6) is $\theta_1(z) = \theta(z) \exp\left(\frac{\pi}{2}\,{}^t z (\operatorname{Im} \tau)^{-1} z\right)$, and that the corresponding "K-function" is $K(m + \tau n) = {}^t mn/2$.

A.6. Jacobians over $\mathbb{C}$

In this section we will sketch the construction of the Jacobian of a compact Riemann surface. The Jacobian will be a complex torus carrying a nondegenerate Riemann form, that is, an abelian variety. The complex-analytic theory in this section parallels the algebro-geometric theory that we will develop in Section A.8. The Jacobian is one of the central tools in studying curves and is the reason why abelian varieties enter into the picture. Indeed, the theory of Jacobians plays an essential role in the proof of Mordell's conjecture. Combining the theory of Jacobians with Theorem A.5.0.1 from the previous section, we also get a proof that all compact Riemann surfaces can be embedded into projective space.

This section is essentially historical and contains few proofs. There are many sources for readers wishing to pursue this very beautiful subject, for example Bost [1], Griffiths and Harris [1], Gunning [1], Lang [4], or the original works of Abel, Jacobi, Riemann, and others.

A.6.1. Abelian Integrals

The theory of abelian varieties arose during the nineteenth century through the attempt to compute or describe integrals of the form $\int R\big(t, \sqrt{P(t)}\,\big)dt$, where R is a rational function and P is a polynomial. More generally, one may consider integrals $\int R(t,s)\,dt$, where s and t satisfy an algebraic relation $P(s,t) = 0$. Such integrals eventually came to be called *abelian integrals.*

As a first example, consider the integral $u = \int_0^x 1/\sqrt{1-t^2}\,dt$. Every student of calculus knows that $u = \sin^{-1}(x)$, so it is easier to look at the inverse function to u. In other words, we consider the function S satisfying $x = S(u)$, and then we find that S is the sine function. In particular, it has a period $S(u+2\pi) = S(u)$ and it satisfies a differential equation $S(u)^2 + S'(u)^2 = 1$. More precisely, the map $u \to \big(S(u), S'(u)\big)$ gives a parametrization of the curve $x^2+y^2=1$.

We are now going to consider the next nontrivial case. Let $Q(t)$ be a polynomial of degree 3 or 4 with distinct roots, and consider the integral $u = \int_0^x 1/\sqrt{Q(t)}\,dt$. As before, we consider the inverse function, which we will denote by $x = f(u)$. The function f has two $\mathbb{R}$-linearly independent periods, and it satisfies the differential equation $f'(u)^2 = Q\big(f(u)\big)$. The map $u \to \big(f(u), f'(u)\big)$ parametrizes the curve $y^2 = Q(x)$. This curve is called an *elliptic curve* because integrals of this sort arise when one tries to

compute the arc length of an ellipse. For a similar reason, such integrals are called *elliptic integrals*, and the corresponding inverse functions are called *elliptic functions*.

Notice that in both cases, the existence of periods for the inverse function comes from the multivaluedness of the integrals. Carrying on with this analogy, we observe that the trigonometric functions satisfy an addition formula,

$$S(u+v) = S(u)S'(v) - S'(v)S(u),$$

and similarly an elliptic function $f(u)$ as above will satisfy an addition formula

$$f(u+v) = F\big(f(u), f'(u), f(v), f'(v)\big)$$

for some rational function F. When $Q(t) = t^3 + At + B$, we have explicitly computed the rational function F, see (A.4.4).

An important discovery of Abel was that when one considers integrals of the form $u = \int_0^x 1/\sqrt{Q(t)}\,dt$ with $\deg(Q) \geq 5$, then one needs to use additional variables. The number of required variables is the "genus" of the integral. For example, suppose that Q has degree 5 or 6 and has distinct roots. Then we define two functions, each depending on two variables, by

$$u_1 = \int_0^{x_1} \frac{1}{\sqrt{Q(t)}}dt + \int_0^{x_2} \frac{1}{\sqrt{Q(t)}}dt,$$
$$u_2 = \int_0^{x_1} \frac{t}{\sqrt{Q(t)}}dt + \int_0^{x_2} \frac{t}{\sqrt{Q(t)}}dt.$$

Consider "inverse functions" f_1 and f_2 satisfying $x_1 + x_2 = f_1(u_1, u_2)$ and $x_1x_2 = f_2(u_1, u_2)$. Then one can show that f_1 and f_2 have four $\mathbb{R}$-linearly independent periods (in $\mathbb{C}^2$). Further, they satisfy an addition formula in which each $f_i(u_1 + v_1, u_2 + v_2)$ can be expressed as a rational function of f_1, f_2, and their derivatives.

A.6.2. Periods of Riemann Surfaces

We now give the modern formulation of the material discussed in the previous section. Let X be a compact Riemann surface of genus g (i.e., a smooth projective curve over $\mathbb{C}$, as we will see). For any regular 1-form ω on X and any path γ on X, we can compute the integral $\int_\gamma \omega$.

Example A.6.2.1. Let P be a polynomial of degree $2g+2$ without multiple roots, and let X be the Riemann surface obtained by gluing the two affine curves $y^2 = P(x)$ and $v^2 = P^*(u) = u^{2g+2}P(u^{-1})$ using the map $(u, v) = (x^{-1}, yx^{-1-g})$. The set $\{dx/y, x\,dx/y, \ldots, x^{g-1}dx/y\}$ is a basis of regular differentials on X. (See Section A.4.5 and Exercise A.4.2 for further

information about these hyperelliptic curves.) Let $\omega = dx/y$, and let γ be a path on X going from $\left(a, \sqrt{P(a)}\right)$ to $\left(b, \sqrt{P(b)}\right)$. Then the line integral $\int_\gamma \omega$ on the Riemann surface X gives a precise meaning to the multivalued integral $\int_a^b 1/\sqrt{P(t)}\,dt$. Of course, it is the choice of the path γ that has eliminated the indeterminacy.

The dependence of the integral on the path is best formulated in terms of homology. Let $\gamma_1, \ldots, \gamma_{2g}$ be a basis of the homology $H_1(X, \mathbb{Z})$ of X. If γ and γ' are two paths joining the points A and B, then γ followed by the reverse of γ' is a closed path, so it is homologous to $\sum m_i \gamma_i$ for some integers m_i. It follows that for any regular 1-form we have

$$\int_\gamma \omega - \int_{\gamma'} \omega = \sum_{i=1}^{2g} m_i \int_{\gamma_i} \omega.$$

Now let $\omega_1, \ldots, \omega_g$ be a basis of the vector space of regular 1-forms, and let Ω be the $g \times 2g$ matrix with entries

$$\Omega = (\Omega_i^j)_{\substack{1 \le i \le 2g \\ 1 \le j \le g}} = \left(\int_{\gamma_i} \omega_j \right)_{\substack{1 \le i \le 2g \\ 1 \le j \le g}}.$$

We call Ω a *period matrix* of X, and we let L_Ω be the $\mathbb{Z}$-module generated by the columns of Ω. (It will soon be apparent that L_Ω is a lattice.) Choosing a different basis for the homology and the space of 1-forms will give another period matrix $\Omega' = A\Omega M$, where $A \in \mathrm{GL}(g, \mathbb{C})$ and $M \in \mathrm{GL}(2g, \mathbb{Z})$. The following beautiful theorem was discovered by Riemann:

Theorem A.6.2.2. (Riemann's period relations) *Let $\gamma_1, \ldots, \gamma_{2g}$ be a basis for the homology group $H_1(X, \mathbb{Z})$, chosen to satisfy the following intersection property: For each $1 \le i \le g$,*

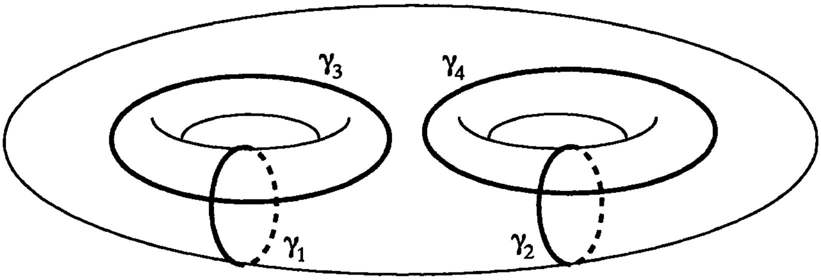
$$\gamma_i \cdot \gamma_j = \begin{cases} 1 & \text{if } j = i + g, \\ 0 & \text{otherwise.} \end{cases}$$

(See Figure A.4 for the case of genus 2.) Then for any nonzero regular 1-forms ω and ω',

$$\sum_{k=1}^{g} \left(\int_{\gamma_k} \omega \int_{\gamma_{g+k}} \omega' - \int_{\gamma_k} \omega' \int_{\gamma_{g+k}} \omega \right) = 0.$$

$$\sqrt{-1} \sum_{k=1}^{g} \left(\overline{\int_{\gamma_{g+k}} \omega} \int_{\gamma_k} \omega - \int_{\gamma_{g+k}} \omega \overline{\int_{\gamma_k} \omega} \right) > 0.$$

PROOF. The proof may be found in Bost [1, III.1.2], Griffiths and Harris [1, page 231], Lang [4, IV.4], or Swinnerton-Dyer [1, I, Theorem 8]. □



A curve of genus 2 with a homology basis
Figure A.4

If we decompose $\Omega = (\Omega_1, \Omega_2)$, where each of Ω_1 and Ω_2 is a $g \times g$ matrix, then Riemann's relations can be written in matrix form as

$$\Omega_1 {}^t\Omega_2 = \Omega_2 {}^t\Omega_1 \qquad \text{and} \qquad -\sqrt{-1}(\overline{\Omega_1}\,{}^t\Omega_2 - \overline{\Omega_2}\,{}^t\Omega_1) > 0.$$

(We write $M > 0$ to indicate that a matrix M is positive definite. That is, ${}^t\overline{Y}MY \geq 0$, with equality if and only if $Y = 0$.) We claim that Ω_1 is invertible. To see this, suppose that ${}^t\Omega_1 Y = 0$. Then

$${}^t\overline{Y}\left(-\sqrt{-1}(\overline{\Omega_1}\,{}^t\Omega_2 - \overline{\Omega_2}\,{}^t\Omega_1)\right) Y = 0,$$

so $Y = 0$. Hence Ω_1 is invertible.

We can thus change our basis of differential forms to transform Ω_1 into the identity matrix and Ω_2 into the matrix $\tau = \Omega_1^{-1}\Omega_2$. In terms of the new period matrix $\Omega = (I, \tau)$, Riemann's relations say that τ is symmetric and that its imaginary part $\mathrm{Im}(\tau)$ is positive definite. The next result is an easy consequence of these observations.

Corollary A.6.2.3. *The column vectors of Ω generate a lattice L_Ω inside $\mathbb{C}^g$.*

PROOF. Using the new basis, the lattice has the form $L_\Omega = \mathbb{Z}^g + \tau\mathbb{Z}^g$. □

A.6.3. The Jacobian of a Riemann Surface

We retain the setting and notation from the previous section.

Definition A.6.3.1. The *Jacobian* of a Riemann surface X is the complex torus $\mathrm{Jac}(X) = \mathbb{C}^g/L_\Omega$, where L_Ω is the lattice generated by the columns of the period matrix Ω. (See Corollary A.6.2.3.)

We next give a more intrinsic formulation. Let V^* denote the dual vector space of a complex vector space V, let $H^0(X,\Omega_X^1)$ be the vector space of regular differentials on X, and let $H_1(X,\mathbb{Z})$ be the homology group of X. We can identify $H_1(X,\mathbb{Z})$ as a lattice in $H^0(X,\Omega_X^1)^*$ via the map

$$H_1(X,\mathbb{Z}) \longrightarrow H^0(X,\Omega_X^1)^*, \qquad \gamma \longmapsto \left(\omega \mapsto \int_\gamma \omega\right).$$

Then the Jacobian of X is equal to

$$\mathrm{Jac}(X) = H^0(X,\Omega_X^1)^*/H_1(X,\mathbb{Z}).$$

We now explicitly construct a Riemann form with respect to the lattice L_Ω. We may assume the lattice to be normalized, $L_\Omega = \mathbb{Z}^g + \tau\mathbb{Z}^g$. Then the form is easy to write down, namely $H(z,w) = {}^t z \,\mathrm{Im}(\tau)^{-1}\overline{w}$. This form is positive definite from Riemann's relations, and if k,ℓ,m,n are vectors with integer coordinates, then $\mathrm{Im}\, H(m+\tau n, k+\tau\ell) = {}^t m\ell - {}^t nk$ is an integer. Hence H is a Riemann form. It follows from Theorem A.5.0.1 that the Jacobian $\mathbb{C}^g/L_\Omega$ is a projective variety. This implies that X is also projective, since X can be embedded in its Jacobian. More precisely, for each fixed basepoint $a \in X$ we define a holomorphic map

$$\Phi_a : X \longrightarrow \mathrm{Jac}(X) = \mathbb{C}^g/L_\Omega, \qquad b \longrightarrow \left(\int_a^b \omega_1, \ldots, \int_a^b \omega_g\right) \bmod L_\Omega.$$

The map Φ_a is called the *Jacobian embedding of* X. (We will explain below why it is an embedding when $g \geq 1$.)

We observe that up to translation, the map Φ_a is independent of a. Thus $\Phi_{a'}(b) = \Phi_a(b) - \Phi_a(a')$. So if we extend Φ_a linearly to the divisor group, then it will be completely independent of a on the the group of divisors of degree zero. We denote this map by Φ,

$$\Phi : \mathrm{Div}^0(X) \longrightarrow \mathrm{Jac}(X), \qquad \sum n_i(b_i) \longmapsto \sum n_i \Phi_a(b_i).$$

The importance of the map Φ comes from the following celebrated theorem.

Theorem A.6.3.2. (Abel–Jacobi) *The map* $\Phi : \mathrm{Div}^0(X) \longrightarrow \mathrm{Jac}(X)$ *is surjective, and its kernel is exactly the subgroup of principal divisors.*

PROOF. See Griffiths and Harris [1, pages 232–237], Lang [4, IV.2.3], or Bost [1, Corollary II.3.5]. □

Corollary A.6.3.3. *Assume that X has genus $g \geq 1$. Then the map $\Phi_a : X \to \mathrm{Jac}(X)$ is an embedding.*

PROOF. Using the theorem, we can identify $\mathrm{Jac}(X)$ with $\mathrm{Pic}^0(X)$. If $\Phi_a(x) = \Phi_a(y)$, then $(x) \sim (y)$, so $x = y$. (Otherwise, X would be a

rational curve, contrary to assumption.) This shows that Φ_a is injective. Further, we see directly from the definition that $\Phi_a^*(dz_i) = \omega_i$, hence Φ_a is an embedding. □

One consequence of (A.6.3.3) is that if X has genus one, then X is isomorphic to its Jacobian. Further, a divisor $\sum n_i(P_i)$ will be principal if and only if $\sum n_i = 0$ and $\sum n_i P_i = 0$. Of course, we already know this from the algebraic proof of Theorem A.4.4.2.

Suppose now that X has genus $g \geq 2$. Then the r-fold sum

$$W_r(X) = \Phi_a(X) + \cdots + \Phi_a(X) = \{\Phi_a(x_1) + \cdots + \Phi_a(x_r) \mid x_1 \ldots, x_r \in X\}$$

is a subvariety of $\mathrm{Jac}(X)$ of dimension $\min(r, g)$. In particular, $\Theta = W_{g-1}(X)$ is a divisor on J. Note that up to translation, Θ is independent of the choice of basepoint $a \in X$. It can be shown that the Riemann form associated to this divisor is precisely the Riemann form we constructed. In particular, the divisor Θ is ample, since the corresponding form is nondegenerate (Theorem A.5.2.7). This is a consequence of the following more precise result.

Theorem A.6.3.4. (Riemann) *Let $L_\Omega = \mathbb{Z}^g + \tau\mathbb{Z}^g$ be the normalized lattice of periods of a Riemann surface X. Then the Riemann theta function*

$$\theta(z) = \theta(z, \tau) = \sum_{m \in \mathbb{Z}^g} \exp(\pi i\, {}^t m \tau m + 2\pi i\, {}^t m z)$$

has associated Riemann form

$$H(z, w) = {}^t z\, \mathrm{Im}(\tau)^{-1} \overline{w}.$$

The divisor associated to this Riemann form is a translate of Θ.

PROOF. The first statement is essentially the computation of the functional equation of θ done in (A.5.2.1(b)). For the second statement, see Bost [1, theorem III.5.1], Griffiths and Harris [1, page 338], or Mumford [5, II, Corollary 3.6]. □

Remark A.6.3.5. The importance of the Riemann form and its associated divisor Θ comes from the fact that the curve X is characterized (up to isomorphism) by the pair $(\mathrm{Jac}(X), \Theta)$. For $g = 2$ this is immediate from Theorem A.6.3.4, since in this case Θ itself is isomorphic to X. The general case is called Torelli's theorem.

A.6.4. Albanese Varieties

In this brief section we explain how the construction of the Jacobian extends to varieties of higher dimension. Let X be a smooth projective variety, and let $H^0(X, \Omega_X^1)$ be the vector space of holomorphic 1-forms on X. Just as for curves, we can embed the first homology group $H_1(X, \mathbb{Z})$ into the dual of $H^0(X, \Omega_X^1)$ via the map

$$H_1(X, \mathbb{Z}) \longrightarrow H^0(X, \Omega_X^1)^*, \qquad \gamma \longmapsto \left(\omega \mapsto \int_\gamma \omega\right).$$

The *Albanese variety of* X, denoted by $\mathrm{Alb}(X)$, is defined to be the torus

$$\mathrm{Alb}(X) = H^0(X, \Omega_X^1)^*/H_1(X, \mathbb{Z}).$$

It can be shown that $H_1(X, \mathbb{Z})$ is a lattice in $H^0(X, \Omega_X^1)^*$, so $\mathrm{Alb}(X)$ is indeed a torus, and in fact, it is an abelian variety (see Weil [5]). Further, there is a map $\Phi_a : X \to \mathrm{Alb}(X)$ defined in exactly the same way as for curves. This construction, although sometimes useful, is not as powerful as the corresponding result for curves, because in general Φ_a will not be injective. More precisely, the map Φ_a is the maximal map of X into an abelian variety in the sense that any other map to an abelian variety will factor through Φ_a. It follows that a smooth projective variety admits a nonconstant map to an abelian variety if and only if it possesses a nonzero regular 1-form. For example, a smooth hypersurface $X \subset \mathbb{P}^n$ with $n \geq 3$ has $\mathrm{Alb}(X) = \{0\}$.

Exercises

A.6.1. Let $P(s,t) \in \mathbb{C}[s,t]$ be a polynomial such that the curve $P(s,t) = 0$ is a rational plane curve, and let $R(s,t) \in \mathbb{C}(s,t)$ be a function on this curve. Prove that the integral $\int R(s,t)\,dt$ can be transformed by a change of variables into an integral $\int F(u)\,du$ for some $F(u) \in \mathbb{C}(u)$. Note that this last integral can be explicitly computed using standard techniques (i.e., partial fractions).

A.6.2. Consider the ellipse $(x/a)^2 + (y/b)^2 = 1$ with $a \geq b > 0$, and let $c^2 = a^2 - b^2$ with $c > 0$. Show that the computation of the arc length of this ellipse leads to the computation of an integral of the form

$$\int \frac{c^2u^2 + b^2}{\sqrt{(1-u^2)(c^2u^2+b^2)}}\,du.$$

Show that the curve $w^2 = (1-u^2)(c^2u^2+b^2)$ is an elliptic curve except when $c = 0$, that is, when the ellipse is a circle.

A.6.3. Let Ω be a period matrix and let J be the skew symmetric matrix whose coefficients are the intersection indices $\gamma_i \cdot \gamma_j$. Prove that Riemann's period relations can be written as

$$\Omega J^{-1}\,{}^t\Omega = 0 \qquad \text{and} \qquad \sqrt{-1}\Omega J^{-1}\,{}^t\overline{\Omega} > 0.$$

A.6.4. Let X be a Riemann surface.

(a) Show that the Jacobian embedding $\Phi_a : X \to \mathrm{Jac}(X)$ induces an isomorphism between $H^0(X, \Omega^1)$ and $H^0(\mathrm{Jac}(X), \Omega^1)$.

(b) Prove that $H^0(\mathrm{Jac}(X), \Omega^1)$ is canonically isomorphic to the tangent space at the origin, $T_0(\mathrm{Jac}(C))$.

A.6.5. The purpose of this exercise is to develop a method to compute the period matrix $\Omega = (I, \tau)$ of certain special curves. In particular, we will compute the period matrix for the curve $C : y^2 = x^6 - 1$. We will need to assume some nontrivial topological facts about Riemann surfaces; see, for example, Lange and Birkenhake [1, Chapter 11, Section 7].

(a) Let $\gamma_1, \ldots, \gamma_{2g}$ be a basis of $H_1(C, \mathbb{Z})$ as described in (A.6.2.2), and let $\omega_1, \ldots, \omega_g$ be a basis of $H^0(C, \Omega_C^1)$. Any automorphism $f \in \mathrm{Aut}(C)$ acts on differential forms and on homology. Let

$${}^tM = {}^t\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Sp}(2g, \mathbb{Z}) \qquad \text{and} \qquad U \in \mathrm{Mat}(g \times g, \mathbb{C})$$

be respectively the matrices giving the actions of f on homology and on differential forms. Prove that

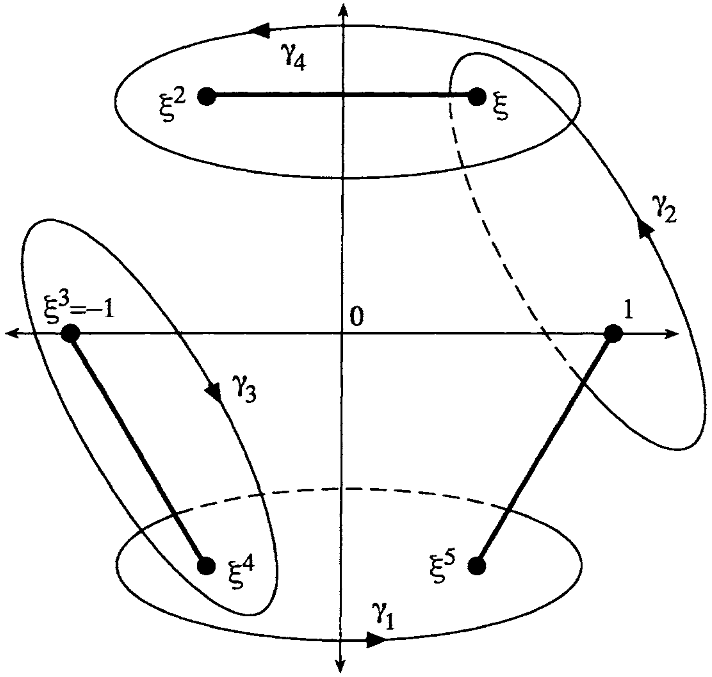
$$\tau = (a\tau + b)(c\tau + d)^{-1} \qquad \text{and} \qquad U = {}^t(c\tau + d).$$

(b) Let $\xi = \exp(2\pi i/6)$ be a primitive sixth root of unity. Consider the curve $C : y^2 = x^6 - 1$, the automorphism $f(x, y) = (\xi x, -y)$, and the differentials $(\omega_1, \omega_2) = (dx/y, x\,dx/y)$. Prove that the corresponding transformation matrix U' described in (a) is given by

$$U' = \begin{pmatrix} -\xi & 0 \\ 0 & -\xi^2 \end{pmatrix}.$$

Deduce that $\mathrm{Trace}(U) = -i\sqrt{3}$ and $\det(U) = -1$.

(c) Topologically, C may be constructed as a two-sheeted covering of $\mathbb{P}^1$ ramified above the six points $1, \xi, \ldots, \xi^5$. Thus it can be represented as two sheets (i.e., two copies of the complex plane) glued along three cuts ("schnitten") joining these ramification points as illustrated in Figure A.5 (cf. Lange-Birkenhake [1, page 346]). The figure also shows four loops $\gamma_1, \ldots, \gamma_4$ that form a basis for $H_1(C, \mathbb{Z})$, where the dashed lines indicate the parts of the paths lying on the lower sheet.



The Riemann surface $y^2 = x^6 - 1$ with cuts and loops
Figure A.5

Show that the transformation matrix of f described in (a) is given by

$$ {}^tM = \left(\begin{array}{cc|cc} 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & -1 \\ \hline 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{array} \right). $$

(*Hint.* Compute intersections $f(\gamma_i) \cdot \gamma_j$.)

(d) Piece together the information gathered above to show that C has a period matrix given by

$$ \Omega = \begin{pmatrix} 2i/\sqrt{3} & i/\sqrt{3} & 1 & 0 \\ i/\sqrt{3} & 2i/\sqrt{3} & 0 & 1 \end{pmatrix}. $$

(e) Let $\rho = (-1 + i\sqrt{3})/2$, and let E be the elliptic curve $E = \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\rho)$. Prove that $\text{Jac}(C)$ is isogenous to the product $E \times E$.

A.7. Abelian Varieties over Arbitrary Fields

In this section we will give a purely algebraic description of the geometry of abelian varieties. In particular, we allow the field of definition to have positive characteristic. Our main tools will be projective geometry and the addition law, in particular the translation maps $t_a : A \to A$ and multiplication maps $[n] : A \to A$.

A.7.1. Generalities

We start by recalling that any algebraic group, such as an abelian variety, is automatically smooth. This is true because it has at least one smooth point, hence a smooth open subset U, and then the translation maps can be used to cover the algebraic group with copies of U. On the other hand, we have defined an abelian variety to be a projective algebraic group, and it is not at all obvious from this definition that the group law must be commutative. In order to prove this fact, we will need the following basic lemma from projective geometry.

Lemma A.7.1.1. (Rigidity lemma) *Let X be a projective variety, let Y and Z be any varieties, and let $f : X \times Y \to Z$ be a morphism. Suppose that there is a point $y_0 \in Y$ such that f is constant on $X \times \{y_0\}$. Then f is constant on every slice $X \times \{y\}$.*

If f is also constant on some slice $\{x_0\} \times Y$, then f is a constant function on all of $X \times Y$.

PROOF. The variety X is projective, and projective varieties are proper (Hartshorne [1, Theorem II.4.9]), so the projection map $p : X \times Y \to Y$ is closed. Hence if U is an affine neighborhood of $z_0 = f(x, y_0)$, then the set $W = p\big(f^{-1}(Z \smallsetminus U)\big)$ is closed in Y. By the hypothesis, $y_0 \notin W$; hence $Y \smallsetminus W$ is a dense open subset of Y. For any $y \notin W$, the projective variety $f\big(X \times \{y\}\big)$ is contained in the affine open set U, hence is reduced to a point. This completes the proof of the first statement of the lemma, and the second statement is clear. □

Notice that the hypothesis that X is projective is crucial. For example, the map $\mathbb{A}^1 \times \mathbb{A}^1 \to \mathbb{A}^1$ given by $(x, y) \mapsto xy$ is constant on the slices $\mathbb{A}^1 \times \{0\}$ and $\{0\} \times \mathbb{A}^1$, but it is certainly not the constant map.

Corollary A.7.1.2. *Let $\phi : A \to B$ be a morphism between two abelian varieties. Then ϕ is the composition of a translation and a homomorphism.*

PROOF. Let e_A and e_B be the identity elements of A and B, respectively. Composing ϕ with a translation, we may assume that $\phi(e_A) = e_B$. Since we

do not yet know that the group laws are commutative, we will temporarily write them multiplicatively. Consider the map

$$f : A \times A \longrightarrow B, \qquad f(x,y) = \phi(xy)\phi(x)^{-1}\phi(y)^{-1}.$$

It is clear that $f\big(\{e_A\}\times A\big) = \{e_B\}$ and $f\big(A\times\{e_A\}\big) = \{e_B\}$, so the rigidity lemma (A.7.1.1) says that f is a constant. Hence $f(x,y) = f(e_A, e_A) = e_B$, which means precisely that ϕ is a homomorphism. □

Notice that Corollary A.7.1.2 is analogous to Lemma A.5.1.1.

Lemma A.7.1.3. *An abelian variety is a commutative algebraic group.*

PROOF. Corollary A.7.1.2 tells us that the inversion morphism $i : A \to A$, $i(x) = x^{-1}$, must be a homomorphism. Hence $i(xy) = i(x)i(y)$, so A is commutative. (See Exercise A.7.3 for another proof closer to the analytic one.) □

We now know that the group law on an abelian variety is commutative, so we will henceforth write the group law additively.

Rational maps from varieties to group varieties have the following important property.

Lemma A.7.1.4. (Weil) *A rational map from a smooth variety into an algebraic group either extends to a morphism or is undefined on a set of pure codimension one.*

PROOF. See Weil [2, 3], Artin [1, Proposition 1.3], or Silverman [2, IV.6.2]. □

Corollary A.7.1.5. *A rational map from a smooth variety into an abelian variety extends to a morphism.*

PROOF. Since an abelian variety is projective, the set of points where the map is not defined has codimension at least two (Theorem A.1.4.4). Then we can use Lemma A.7.1.4 to conclude that the map extends to a morphism. □

For example, Corollary A.7.1.5 implies that a rational map from $\mathbb{P}^n$ to an abelian variety is constant (see Exercise A.7.4). Corollary A.7.1.5 is a powerful tool in analyzing maps to an abelian variety. It is complemented by the next proposition, which describes maps from an abelian variety.

Proposition A.7.1.6. *Let A be an abelian variety, and let $f : A \to Y$ be a morphism. Then there is an abelian subvariety B of A such that for*

any $x \in A$, the connected component of $f^{-1}(f(x))$ containing x is equal to $B+x$.

PROOF. Let C_x be the connected component of $f^{-1}\{f(x)\}$ containing x, and let $B = C_0$. We consider the map

$$\phi : A \times C_x \longrightarrow Y, \qquad (a,u) \longmapsto f(a+u).$$

We note that $\phi(\{0\} \times C_x)$ is a point. The rigidity lemma (A.7.1.1) implies that $\phi(\{a\} \times C_x)$ is a point for any a. Equivalently, $f(a+C_x) = f(a+x)$. But $a-x+C_x$ is connected and contains a, so we see that $a-x+C_x \subset C_a$. By symmetry, we must have equality. (Note that the rigidity lemma applies to each irreducible component.) Putting $a=0$ gives $C_x = x+B$, so it remains to show that B is a subgroup. If $b \in B$, then $C_{-b} = -b+B$, so $0 \in C_{-b}$. Hence $C_{-b} = B$, or equivalently, $-b+B \subset B$, which shows that B is a subgroup. □

A.7.2. Divisors and the Theorem of the Cube

In this section we will study divisors on abelian varieties. Since abelian varieties are smooth, we do not need to worry about distinguishing between Cartier and Weil divisors, so we will write $\mathrm{Pic}(A)$ for the divisor class group of A and we will use $\sim$ to denote linear equivalence. We are especially interested in divisor relations that reflect the group structure, and we will also want to derive a criterion for ampleness.

In order to state our first result, we need to define various projection–summation maps. Thus for any subset I of $\{1,2,3\}$, we define a map

$$s_I : A \times A \times A \longrightarrow A, \qquad s_I(x_1,x_2,x_3) = \sum_{i \in I} x_i.$$

For example, $s_{13}(x_1,x_2,x_3) = x_1 + x_3$ and $s_2(x_1,x_2,x_3) = x_2$. We are now ready for the following fundamental theorem.

Theorem A.7.2.1. (Theorem of the cube on abelian varieties) *Let A be an abelian variety. Then for every divisor $D \in \mathrm{Div}(A)$, the following divisor class relation holds in $A \times A \times A$:*

$$s_{123}^*D - s_{12}^*D - s_{13}^*D - s_{23}^*D + s_1^*D + s_2^*D + s_3^*D \sim 0.$$

PROOF. To ease notation, we will let

$$\begin{aligned}\operatorname{cube}(D) &= s_{123}^*D - s_{12}^*D - s_{13}^*D - s_{23}^*D + s_1^*D + s_2^*D + s_3^*D \\ &= -\sum_{I\subset\{1,2,3\}} (-1)^{\#I} s_I^*(D)\end{aligned}$$

be the divisor sum we are studying. We start by giving a proof when the ground field is $\mathbb{C}$. It is enough to prove the theorem for effective divisors D, which means that D is the divisor of some theta function θ. We define a function on $A \times A \times A$ by

$$F(z_1, z_2, z_3) = \frac{\theta(z_1 + z_2 + z_3)\theta(z_1)\theta(z_2)\theta(z_3)}{\theta(z_1 + z_2)\theta(z_1 + z_3)\theta(z_2 + z_3)}.$$

It is clear that $\operatorname{div}(F) = \operatorname{cube}(D)$. Further, using the functional equation of the function θ, we find that the automorphy factor of F is trivial. In other words, F is a meromorphic function on $A \times A \times A$ whose divisor is $\operatorname{cube}(D)$, so $\operatorname{cube}(D) \sim 0$.

This proves the theorem over $\mathbb{C}$, and so by the Lefschetz principle over any field of characteristic zero. In the general case, we can deduce Theorem A.7.2.1 from the following more general result, which also explains the word "cube" in the name of the theorem.

Theorem A.7.2.2. (Theorem of the cube) *Let X, Y, and Z be projective varieties, and let $(x_0, y_0, z_0) \in X \times Y \times Z$. Let D be a divisor on $X \times Y \times Z$ whose linear equivalence class becomes trivial when restricted to each of the three slices*

$$X \times Y \times \{z_0\}, \qquad X \times \{y_0\} \times Z, \qquad \text{and} \qquad \{x_0\} \times Y \times Z.$$

Then D is linearly equivalent to zero on $X \times Y \times Z$.

PROOF. See Mumford [2, II.6]. □

We now explain how Theorem A.7.2.2 can be used to prove Theorem A.7.2.1. Let i be the injection

$$i : A \times A \to A \times A \times A, \qquad i(x_1, x_2) = (x_1, x_2, 0).$$

We apply Theorem A.7.2.2 with $X = Y = Z = A$ and $x_0 = y_0 = z_0 = 0$. Theorem A.7.2.2 (and symmetry) say that it is enough to show that $i^*\bigl(\operatorname{cube}(D)\bigr) = 0$ in $\operatorname{Pic}(A \times A)$. To do this, we compute

$$\begin{aligned}s_{123} \circ i(x_1, x_2) &= x_1 + x_2 = s_{12} \circ i(x_1, x_2), \\ s_{23} \circ i(x_1, x_2) &= x_2 = s_2 \circ i(x_1, x_2), \\ s_{13} \circ i(x_1, x_2) &= x_1 = s_1 \circ i(x_1, x_2), \\ s_3 \circ i(x_1, x_2) &= 0.\end{aligned}$$

Hence all of the terms in the sum

$$i^*(\mathrm{cube}(D)) = \sum_{I\subset\{1,2,3\}} (-1)^{\#(I)}(s_I \circ i)^*(D)$$

cancel. This completes the proof of Theorem A.7.2.2. □

Let p_{23}, p_{13}, p_{12} be the projections from $X\times Y\times Z$ onto $Y \times Z$, $X \times Z$, and $X \times Y$, respectively. Then the theorem of the cube can be rephrased as saying that the map

$$\begin{aligned} \operatorname{Pic}(Y \times Z) \times \operatorname{Pic}(X \times Z) \times \operatorname{Pic}(X \times Y) &\longrightarrow \operatorname{Pic}(X \times Y \times Z), \\ (c_1, c_2, c_3) &\longmapsto p_{23}^* c_1 + p_{13}^* c_2 + p_{12}^* c_3, \end{aligned}$$

is surjective, or in fancier language, that the functor Pic is quadratic. (See Mumford [2, II.6] or Serre [3].) The theorem is not true with only two factors. For example, let C be a curve of genus $g \geq 1$, and let D be the divisor $\Delta - C\times\{P\}-\{P\}\times C$ on the product $C \times C$. Then the restrictions of D to the two slices $\{P\} \times C$ and $C \times \{P\}$ are trivial, but the restriction of D to $C \times \{Q\}$ is (the class of) $(Q) - (P)$, which is not trivial unless $P = Q$ or C is rational. The correct statement for two factors, which is an intermediate step in the proof of the theorem of the cube, is called the seesaw principle.

Lemma A.7.2.3. (Seesaw principle) *Let X and Y be two varieties, let $c \in \operatorname{Pic}(X \times Y)$, and define maps $i_x(y) = (x, y)$ and $p_1(x, y) = x$.*

(i) *If $i_x^*(c) = 0$ in $\operatorname{Pic}(Y)$ for all $x \in X$, then there exists a class $c' \in \operatorname{Pic}(X)$ such that $c = p_1^*(c')$.*

(ii) *If furthermore c is trivial when restricted to some slice $X \times \{y_0\}$, then $c = 0$ in $\operatorname{Pic}(X \times Y)$.*

PROOF. (Sketch) Let D be a divisor in the class c. For all x in some open subset U of X, we have $i_x^*(D) = \operatorname{div}(f_x)$. Set $g(x, y) = f_x(y)$ and $D' = D - (g)$. Then, possibly after shrinking U, we find that $i_x^*(D') = 0$ for all $x \in U$. Hence the support of D' is concentrated on $(X \setminus U) \times Y$, and thus D' has the form $D'' \times Y = p_1^* D''$. The second statement is clear. □

We deduce several important corollaries from the theorem of the cube.

Corollary A.7.2.4. *Let A be an abelian variety, let V be an arbitrary variety, and let $f, g, h : V \to A$ be three morphisms from V to A. Then for any divisor $D \in \operatorname{Div}(A)$,*

$$(f+g+h)^*D-(f+g)^*D-(f+h)^*D-(h+g)^*D+f^*D+g^*D+h^*D \sim 0.$$

Note that this is a linear equivalence on $V \times V \times V$, where for example the map $(f+h) : V \times V \times V \to A$ is given by $(f+h)(x,y,z) = f(x) + h(z)$.

PROOF. Let cube(D) be the divisor described in Theorem A.7.2.1. Then the divisor we are analyzing is the pullback of the divisor cube(D) by the map $(f,g,h) : V \times V \times V \to A \times A \times A$. But cube$(D) \sim 0$ from (A.7.2.1), so we are done. □

Corollary A.7.2.4 implies that for any divisor class $c \in \mathrm{Pic}(A)$, the map

$$\mathrm{Mor}(V,A) \times \mathrm{Mor}(V,A) \longrightarrow \mathrm{Pic}(V \times V), \qquad (f,g) \longmapsto (f+g)^*c - f^*c - g^*c,$$

is bilinear. The next corollary gives a quadratic property.

Corollary A.7.2.5. (Mumford's formula) *Let D be a divisor on an abelian variety A, and let $[n] : A \to A$ be the multiplication-by-n map. Then*

$$[n]^*(D) \sim \left(\frac{n^2+n}{2}\right) D + \left(\frac{n^2-n}{2}\right)[-1]^*(D).$$

In particular,

$$[n]^*(D) \sim \begin{cases} n^2 D & \text{if } D \text{ is symmetric } ([-1]^*D \sim D), \\ nD & \text{if } D \text{ is antisymmetric } ([-1]^*D \sim -D). \end{cases}$$

PROOF. The formula is trivially true for $n = -1$, $n = 0$, and $n = 1$. Next we apply Corollary A.7.2.4 with $f = [n]$, $g = [1]$ and $h = [-1]$ to obtain

$$[n+1]^*D + [n-1]^*D - 2[n]^*D \sim D + [-1]^*D.$$

Now an easy induction, both upwards and downwards from $n = 0$, gives the desired result. Or one can use the following elementary lemma. □

Lemma A.7.2.6. *Let G be an abelian group, and let $f : \mathbb{Z} \to G$ be a map with the property that $f(n+1) - 2f(n) + f(n-1)$ is constant. Then*

$$f(n) = \frac{n^2+n}{2} f(1) + \frac{n^2-n}{2} f(-1) - (n^2-1) f(0).$$

PROOF. Any quadratic function $g(n) = an^2 + bn + c$ has the property that $g(n+1) - 2g(n) + g(n-1)$ is constant. In particular, this is true of the function $g(n) = \frac{1}{2}(n^2+n)f(1) + \frac{1}{2}(n^2-n)f(-1) - (n^2-1)f(0)$. On the other hand, any function with this property is completely determined by its values at $n = -1, 0, 1$. Since $g(n) = f(n)$ for $n = -1, 0, 1$, it follows that $g(n) = f(n)$ for all n. □

We now can give an algebraic proof that the kernel of multiplication by n on an abelian variety of dimension g is isomorphic to $(\mathbb{Z}/n\mathbb{Z})^{2g}$, provided that n is relatively prime to the characteristic of the base field. Notice that this fact is obvious for complex tori.

Theorem A.7.2.7. *Let A be an abelian variety of dimension g over an algebraically closed field k of characteristic $p \geq 0$.*
(i) *The multiplication-by-n map $[n] : A \to A$ is a degree n^{2g} isogeny.*
(ii) *Assume either that $p = 0$ or that $p \nmid n$. Then*

$$A[n] = \ker[n] \cong (\mathbb{Z}/n\mathbb{Z})^{2g}.$$

(iii) *If $p > 0$, then $A[p^t] \cong (\mathbb{Z}/p^t\mathbb{Z})^r$ for some integer $0 \leq r \leq g$.*

PROOF. (i) The addition map $\mu : A \times A \to A$ on the abelian variety A induces a map $\mu_* : T_0(A) \times T_0(A) \to T_0(A)$ on the tangent space of $A \times A$ at $(0,0)$, and it is not hard to see that μ_* is simply the addition map on the tangent space. It follows by induction that $[n]_*$ is multiplication by n on the tangent space. If $p = 0$ or if $p \nmid n$, then $[n]_*$ is an isomorphism on $T_0(A)$. Therefore, $\dim([n]A) = \dim(A)$, which shows that $[n]$ is surjective and hence is an isogeny. If $p|n$, then $[n]$ is still an isogeny, but since we will not need to use this fact, we will refer the reader to Mumford [2, II.6, page 64] for the proof.

We will use the following lemma to compute the degree of $[n]$.

Lemma A.7.2.8. *Let A be an abelian variety of dimension g over an algebraically closed field k of characteristic $p \geq 0$, and let $\phi : A \to A$ be an isogeny.*
(a) *Let $D_1, \ldots, D_g \in \mathrm{Div}(A)$. Then*

$$(\phi^* D_1, \ldots, \phi^* D_g)_A = \deg(\phi)(D_1, \ldots, D_g)_A.$$

(See Section A.2.3 for a general discussion of intersection indices.)
(b) *If $D \in \mathrm{Div}(A)$ is ample, then $D^g = (D, D, \ldots, D)_A > 0$.*

PROOF. (a) This is a special case of Theorem A.2.3.2.
(b) Replacing D with a multiple, we may assume that D is very ample and use it to define an embedding $F : A \to \mathbb{P}^m$. Since A has dimension g, we can choose hyperplanes $H_1, \ldots, H_g$ such that the intersection $F(A) \cap H_1 \cap \cdots \cap H_g$ is finite, say consisting of N points. Then $D^g = (F^* H_1, \ldots, F^* H_g) \geq N > 0$. □

We now resume the proof of Theorem A.7.2.7. Let $D \in \mathrm{Div}(A)$ be an ample symmetric divisor. (For example, let D' be very ample and take $D = D' + [-1]^* D'$.) Then

$$\begin{aligned} \deg([n]) D^g &= ([n]^* D)^g && \text{(from Lemma A.7.2.8(a))}, \\ &= (n^2 D)^g && \text{(from Corollary A.7.2.5)}, \\ &= n^{2g} D^g && \text{(by linearity of intersection index)}. \end{aligned}$$

But we also know from Lemma A.7.2.8(b) that $D^g > 0$, so we conclude that $\deg([n]) = n^{2g}$. This completes the proof of part (i) of (A.7.2.7).

Suppose now that $p = \mathrm{char}(k) = 0$ or that $p \nmid n$. Then the isogeny $[n]$ is separable, since its degree is prime to p, so its kernel has order equal to the degree. In other words, $\#A[n] = n^{2g}$. Further, this formula is true for every such integer n. The following elementary lemma implies that $A[n] \cong (\mathbb{Z}/n\mathbb{Z})^{2g}$, which will complete the proof of part (ii) of (A.7.2.7). Finally, since we will not need part (iii), we refer the reader to Mumford [2, II.6, page 64] for its proof.

Lemma. *Let A be a finite abelian group of order n^r, and suppose that for every integer $m|n^r$, the m-torsion subgroup $A[m]$ satisfies $\#A[m] = m^r$. Then $A \cong (\mathbb{Z}/n\mathbb{Z})^r$.*

PROOF. For any integer $d \geq 2$, let $C(d)$ denote a cyclic group of order d. The structure theorem for finite abelian groups says that there are integers $d_1|d_2|\cdots|d_s$ such that $A \cong C(d_1) \oplus \cdots \oplus C(d_s)$.

First we observe that d_s kills A, so $n^r = \#A = \#A[d_s] = d_s^r$. Thus $d_s = n$. Next we note that d_1 divides each d_i, so $C(d_i)[d_1] \cong C(d_1)$, and hence $d_1^r = \#A[d_1] = d_1^s$. Therefore, $r = s$. Now, for each i, let $e_i = d_r/d_i = n/d_i$, so the e_i's are integers. Then

$$n^r = \#A = d_1 d_2 \cdots d_r = n^r/(e_1 e_2 \cdots e_r).$$

It follows that $e_1 = \cdots = e_r = 1$, so $d_i = n$ for all n. Hence $A \cong C(n)^r$, which completes the proof of the lemma. □

Having studied the effect of multiplication-by-n on divisors, we next describe the action of the translation maps.

Theorem A.7.2.9. (Theorem of the square) *Let A be an abelian variety, and for each $a \in A$, let $t_a : A \to A$ be the translation-by-a map $t_a(x) = x + a$. Then*

$$t_{a+b}^*(D) + D \sim t_a^*(D) + t_b^*(D) \qquad \text{for all } D \in \mathrm{Div}(A) \text{ and } a, b \in A.$$

In other words, for any divisor class $c \in \mathrm{Pic}(A)$, the map

$$\Phi_c : A \longrightarrow \mathrm{Pic}(A), \qquad a \longmapsto t_a^*(c) - c,$$

is a group homomorphism.

PROOF. We just need to apply Corollary A.7.2.4 with the maps $f(x) = x$, $g(x) = a$, and $h(x) = b$. □

In this section we have used the theorem of the cube as the cornerstone of our theory, but we want to mention that it is also possible to start with the theorem of the square and deduce the theorem of the cube (see Exercise A.7.5). Notice that over $\mathbb{C}$, the theorem of the square is an immediate consequence of the fact that the automorphy factor of a theta function is linear.

Notation. For any divisor $D \in \operatorname{Div}(A)$, we let

$$\Phi_D : A \longrightarrow \operatorname{Pic}(A), \qquad a \longmapsto \operatorname{class}\big(t_a^*(D) - D\big),$$

be the homomorphism described in (A.7.2.9), and we let $K(D) = \ker(\Phi_D)$.

The group $K(D)$ can be used to give an ampleness criterion for divisors on abelian varieties.

Theorem A.7.2.10. *Let D be an effective divisor on an abelian variety A. Then the linear system $|2D|$ is base-point free, and the following four conditions are equivalent:*

(i) *D is ample.*
(ii) *The group $K(D) = \{a \in A \mid t_a^*(D) \sim D\}$ is finite.*
(iii) *The stabilizer $G(D) = \{a \in A \mid t_a^*(D) = D\}$ is finite.*
(iv) *The morphism $A \to \mathbb{P}L(2D)$ associated to $2D$ is a finite morphism.*

PROOF. By the theorem of the square, $t_{-x}^*D + t_x^*D \sim 2D$, and clearly a point $y \in A$ cannot be on every translate of D. Hence $2D$ is base-point free.

Let $f : A \to \mathbb{P}L(2D)$ be the morphism associated to $2D$, and let $B(D)$ be the abelian subvariety attached to f by Proposition A.7.1.6. Then $f^{-1}\big(f(a)\big) = B(D)+a$ for every $a \in A$. In particular, $f^{-1}\big(f(a)\big)$ is invariant under translation by any point $b \in B(D)$. It follows that $t_b^* \circ f^* = f^*$ as maps on divisors. In particular, if we take a hyperplane H in $\mathbb{P}L(2D)$ with $f^*H = 2D$, then we find that

$$2D = f^*(H) = t_b^* \circ f^*(H) = t_b^*(2D) = 2t_b^*D.$$

(N.B. This is an equality of divisors, not merely of divisor classes.) Hence $b \in G(D)$, so we have proven that $B(D) \subset G(D)$.

(ii) $\Longrightarrow$ (iii) This is obvious from the trivial inclusion $G(D) \subset K(D)$.

(iii) $\Longrightarrow$ (iv) The map f is finite if and only if its fibers have dimension 0, so if and only if $B(D)$ is finite. Hence the inclusion $B(D) \subset G(D)$ proven above gives the desired result.

(iv) $\Longrightarrow$ (i) This is a special case of Proposition A.3.2.4(ii)

(i) $\Longrightarrow$ (iii) Let $a \in A$ be a point with $a \notin D$, and let $V = a + G(D) \subset A$. We claim that $V \cap D = \emptyset$. To see this, suppose that $x \in V \cap D$. Then $x = a + g$ for some $g \in G(D)$, and so $a = x - g \in D - g = D$, since D is invariant under translation by g. This contradicts the choice of a, so we see that $V \cap D = \emptyset$. But D is effective and ample by assumption, so Exercise A.3.6(b) implies that $\dim(V) = 0$. Therefore, $G(D)$ is finite.

(iv) $\Longrightarrow$ (ii) Let b be in a connected component $K(D)^0$. Then $\Phi_{2D} \circ t_b = L_b \circ \Phi_{2D}$ for some $L_b \in \operatorname{PGL}(\ell(2D))$. The map $b \mapsto L_b$ is from a projective variety to an affine group, hence must be constant. We conclude that Φ_{2D} is constant on $K(D)^0$, which therefore must be a point. □

A.7.3. Dual Abelian Varieties and Poincaré Divisors

The main purpose of this section is to show that $\operatorname{Pic}(A)$ has a "connected component" that is itself an abelian variety. We start by giving one description of this connected component.

Definition. Let A be an abelian variety. The group $\operatorname{Pic}^0(A)$ is the group of translation-invariant divisor classes,

$$\operatorname{Pic}^0(A) = \bigl\{c \in \operatorname{Pic}(A) \bigm| t_a^* c = c \text{ for all } a \in A\bigr\}.$$

Theorem A.7.3.1. *Let A be an abelian variety, let $c \in \operatorname{Pic}(A)$, and let*

$$\Phi_c : A \longrightarrow \operatorname{Pic}(A), \qquad a \longmapsto t_a^* c - c,$$

be the homomorphism described in Theorem A.7.2.9.
(a) *The image of Φ_c lies in $\operatorname{Pic}^0(A)$.*
(b) *If $nc \in \operatorname{Pic}^0(A)$ for some integer $n \neq 0$, then $c \in \operatorname{Pic}^0(A)$.*
(c) *If the divisor class c is ample, then $\Phi_c : A \to \operatorname{Pic}^0(A)$ is surjective and has a finite kernel.*

PROOF. (a) This is clear from the theorem of the square (A.7.2.9),

$$t_b^*\bigl(\Phi_c(a)\bigr) = t_b^*(t_a^* c - c) = t_{a+b}^* c - t_b^* c = t_a^* c - c.$$

(b) It is clear from the definition of Φ_c that $\Phi_{c+c'} = \Phi_c + \Phi_{c'}$, so using (a) and the definition of $\operatorname{Pic}^0(A)$, we can say that there is an exact sequence

$$\begin{array}{ccccccc} 0 & \longrightarrow & \operatorname{Pic}^0(A) & \longrightarrow & \operatorname{Pic}(A) & \longrightarrow & \operatorname{Hom}\bigl(A, \operatorname{Pic}^0(A)\bigr), \\ & & & & c & \longmapsto & \Phi_c. \end{array}$$

Now suppose that $nc \in \operatorname{Pic}^0(A)$. Then for all $a \in A$ we have

$$0 = \Phi_{nc}(a) = (n\Phi_c)(a) = \Phi_c\bigl([n]a\bigr).$$

But A is n-divisible (i.e., $[n](A) = A$), so Φ_c is the zero map. Hence $c \in \operatorname{Pic}^0(A)$.
(c) See Mumford [2, II.8, Theorem 1] or Lang [3, IV.2, Theorem 4] . A proof over $\mathbb{C}$ is sketched in Exercise A.5.5. □

Remark. The *Néron–Severi group of A*, denoted by $\operatorname{NS}(A)$, is the quotient group $\operatorname{NS}(A) = \operatorname{Pic}(A)/\operatorname{Pic}^0(A)$. Theorem A.7.3.1(b) says that $\operatorname{NS}(A)$ has no torsion. We also see that the map $c \mapsto \Phi_c$ induces an injective homomorphism $\operatorname{NS}(A) \hookrightarrow \operatorname{Hom}\bigl(A, \operatorname{Pic}^0(A)\bigr)$.

The divisor classes in $\operatorname{Pic}^0(A)$ can also be characterized as the anti-symmetric or odd classes.

Proposition A.7.3.2. Let $c \in \mathrm{Pic}(A)$. *The following are equivalent:*
(i) $[-1]^*c = -c$.
(ii) $c \in \mathrm{Pic}^0(A)$, *or equivalently,* $K(c) = A$.
(iii) $s_{12}^*c - p_1^*c - p_2^*c = 0$, *where* $s_{12}, p_1, p_2 : A \times A \to A$ *are the usual maps,* $s_{12}(x,y) = x+y$, $p_1(x,y) = x$, *and* $p_2(x,y) = y$.

PROOF. To ease notation, we will write $\Gamma_c = s_{12}^*c - p_1^*c - p_2^*c$. Also, for any $a \in A$, we let $i_a : A \to A \times A$ be the map $i_a(x) = (a,x)$. Notice that $s_{12} \circ i_a(x) = t_a(x)$, $p_1 \circ i_a(x) = a$, and $p_2 \circ i_a(x) = x$. Using these formulas, we obtain the relation

$$i_a^*(\Gamma_c) = i_a^*(s_{12}^*c - p_1^*c - p_2^*c) = t_a^*c - c.$$

(iii) $\Longrightarrow$ (ii) We are given that $\Gamma_c = 0$, so $t_a^*c - c = i_a^*(\Gamma_c) = 0$. Hence $c \in \mathrm{Pic}^0(A)$.
(ii) $\Longrightarrow$ (iii) We are given that $t_a^*c - c = 0$ for every $a \in A$, so $i_a^*(\Gamma_c) = 0$. Further, Γ_c is clearly trivial when restricted to $A \times \{0\}$, so the seesaw principle (A.7.2.3) implies that $\Gamma_c = 0$.
(ii) $\Longrightarrow$ (i) Fix an ample symmetric divisor class $c_0 \in \mathrm{Pic}(A)$. Theorem A.7.3.1(c) says that there is an $a \in A$ such that $c = t_a^*c_0 - c_0$. We use the theorem of the square (A.7.2.9) to calculate

$$[-1]^*c = [-1]^*t_a^*c_0 - [-1]^*c_0 = t_{-a}^*c_0 - c_0 = -t_a^*c_0 + c_0 = -c.$$

(i) $\Longrightarrow$ (ii) For an arbitrary $c \in \mathrm{Pic}(A)$, we claim that $c - [-1]^*c$ is in $\mathrm{Pic}^0(A)$. To verify this, we compute

$$\begin{aligned}
&t_a^*(c - [-1]^*c) - (c - [-1]^*c) \\
&\quad = t_a^*c - [-1]^*t_{-a}^*c - c + [-1]^*c && \text{since } [-1] \circ t_a = t_{-a} \circ [-1] \\
&\quad = (t_a^*c - c) - [-1]^*(t_{-a}^*c - c) \\
&\quad = (t_a^*c - c) - [-1]^*(c - t_a^*c) && \text{theorem of the square (A.7.2.9)} \\
&\quad = c' + [-1]^*c' && \text{where } c' = t_a^*c - c \text{ (Theorem A.7.3.1} \\
& && \text{says that } c' \in \mathrm{Pic}^0(A).) \\
&\quad = 0 && \text{from (ii)} \Longrightarrow \text{(i)}.
\end{aligned}$$

Now suppose that $[-1]^*c = -c$. Then by what we just proved, $2c = c - [-1]^*c \in \mathrm{Pic}^0(A)$. It follows from Theorem A.7.3.1(b) that $c \in \mathrm{Pic}^0(A)$, which completes the proof of the theorem. □

We can give a similar characterization of symmetric or even classes.

Proposition A.7.3.3. Let $c \in \mathrm{Pic}(A)$. *The following are equivalent:*
(i) $[-1]^*c = c$.
(ii) $s_{12}^*c + d_{12}^*c = 2p_1^*c + 2p_2^*c$, *where* s_{12}, p_1, p_2 *are as in (A.7.3.2) and* $d_{12}(x,y) = x - y$.

PROOF. Let $\Gamma_c = s_{12}^*c + d_{12}^*c - 2p_1^*c - 2p_2^*c$, and let $j : A \to A \times A$ be the map $j(x) = (0,x)$. Then

$$s_{12} \circ j(x) = x, \quad d_{12} \circ j(x) = -x, \quad p_1 \circ j(x) = 0, \quad \text{and} \quad p_2 \circ j(x) = x,$$

so we can compute

$$j^*(\Gamma_c) = j^*(s_{12}^*c + d_{12}^*c - 2p_1^*c - 2p_2^*c) = c + [-1]^*c - 0 - 2c = [-1]^*c - c.$$

(ii) $\Longrightarrow$ (i) This implication is clear from the relation $j^*(\Gamma_c) = [-1]^*c - c$.
(i) $\Longrightarrow$ (ii) Let $i_a(x) = (x, a)$ be as in the proof of Proposition A.7.3.2. Then the theorem of the square (A.7.2.9) tells us that

$$i_a^*(\Gamma_c) = i_a^*(s_{12}^*c + d_{12}^*c - 2p_1^*c - 2p_2^*c) = t_a^*c + t_{-a}^*c - 2c = 0.$$

In other words, the divisor class Γ_c is trivial on every slice $A \times \{a\}$. Notice that we have not yet used the assumption that $[-1]^*c = c$. However, if we make that assumption, then we get $j^*(\Gamma_c) = 0$ from above, so Γ_c is also trivial on the slice $\{0\} \times A$. It follows from the seesaw principle (A.7.2.3) that $\Gamma_c = 0$. □

Notice that Propositions A.7.3.2 and A.7.3.3 say that odd divisor classes have a certain linear property and that even divisor classes have a quadratic property.

We now come to the classical, yet astonishing, fact that $\mathrm{Pic}^0(A)$ can be given the structure of an abelian variety. We formalize the correspondence in the following way.

Definition. An abelian variety $\hat{A}$ is called the *dual abelian variety of A* if there exists a divisor class $\mathcal{P}$ on $A \times \hat{A}$ such that the maps

$$\hat{A} \longrightarrow \mathrm{Pic}^0(A), \qquad \hat{a} \longmapsto i_{\hat{a}}^*(\mathcal{P}),$$

and

$$A \longrightarrow \mathrm{Pic}^0(\hat{A}), \qquad a \longmapsto i_a^*(\mathcal{P}),$$

are both bijections. (Here $i_{\hat{a}} : A \to A \times \hat{A}$ is the map $i_{\hat{a}}(a) = (a, \hat{a})$, and $i_a : \hat{A} \to A \times \hat{A}$ is the map $i_a(\hat{a}) = (a, \hat{a})$.) The divisor class $\mathcal{P}$ is called the *Poincaré divisor class.*

Theorem A.7.3.4. *The dual abelian variety $\hat{A}$ exists and together with the Poincaré class $\mathcal{P} \in \mathrm{Pic}(A \times \hat{A})$ is unique up to isomorphism. Further, the Poincaré class $\mathcal{P}$ is even.*

PROOF. We will give the proof below in the case that there is a divisor class $c \in \mathrm{Pic}(A)$ with $K(c) = 0$. In general, one chooses any ample c. Then $K(c)$ is finite (A.7.3.1), and one takes $\hat{A}$ to be the quotient $A/K(c)$. The Poincaré class is constructed by showing that the divisor class $s_{12}^*c - p_1^*c - p_2^*c$ on $A \times A$ descends to the quotient $\hat{A} \times \hat{A}$. For further details, see Mumford [2, II.8]. See also Exercise A.5.6 for a proof over the complex numbers. □

The map $\Phi_c : A \to \mathrm{Pic}^0(A)$ will induce an isogeny $\Phi_c : A \to \hat{A}$, provided that its kernel $K(c)$ is finite, or equivalently by (A.7.2.10), c is

ample. Such an isogeny is called a *polarization.* It is said to be a *principal polarization* if $K(c) = \{0\}$. Thus c gives a principal polarization if the map $\Phi_c : A \to \hat{A}$ is an isomorphism. Not every abelian variety admits a principal polarization, but we will see in the next section that Jacobian varieties come naturally equipped with a principal polarization. If A does admit a principal polarization, then A is its own dual and it is possible to describe the Poincaré divisor quite precisely.

Theorem A.7.3.5. *Suppose that there exists a divisor class $c \in \operatorname{Pic}(A)$ such that $K(c) = 0$, where $K(c) = \{a \in A \mid t_a^* c = c\}$. Then A is its own dual, and*

$$s_{12}^* c - p_1^* c - p_2^* c \in \operatorname{Pic}(A \times A)$$

is a Poincaré divisor class.

PROOF. Let $\mathcal{P} = s_{12}^* c - p_1^* c - p_2^* c$, and for each $y \in A$, let $i_y : A \to A \times A$ be the map $i_y(x) = (x, y)$. The divisor $\mathcal{P}$ is clearly symmetric, so it suffices to show that the map

$$A \longrightarrow \operatorname{Pic}^0(A), \qquad y \longmapsto i_y^* \mathcal{P},$$

is an isomorphism. Notice that

$$s_{12} \circ i_y(x) = x + y = t_y(x), \quad p_1 \circ i_y(x) = x, \quad \text{and} \quad p_2 \circ i_y(x) = y.$$

Using these, we can compute

$$i_y^* \mathcal{P} = i_y^* \circ s_{12}^* c - i_y^* \circ p_1^* c - i_y^* \circ p_2^* c = t_y^* c - c = \Phi_c(y).$$

In other words, the map $y \longmapsto i_y^* \mathcal{P}$ is equal to Φ_c. But our assumption that $K(c) = 0$ means that Φ_c is an isomorphism (A.7.3.1(c)), so $\mathcal{P}$ is a Poincaré divisor. □

We close this section with a brief mention of how some of these constructions generalize to an arbitrary smooth projective variety V. One can define $\operatorname{Pic}^0(V)$ to be the subgroup of $\operatorname{Pic}(V)$ composed of divisor classes algebraically equivalent to zero (see the remark at the end of Section A.2.3). The group $\operatorname{Pic}^0(V)$ can always be given the structure of an abelian variety, which is called the Picard variety of V. The quotient $\operatorname{NS}(V) = \operatorname{Pic}(V)/\operatorname{Pic}^0(V)$ is called the Néron–Severi group of V and is a finitely generated group. For example, if A is an abelian variety over $\mathbb{C}$, then $\operatorname{NS}(A)$ is the group of Riemann forms on A, and if C is a smooth projective curve, then $\operatorname{NS}(C) = \mathbb{Z}$.

There is another abelian variety associated to V, called the *Albanese variety* $\operatorname{Alb}(V)$. (See Section A.6.4 for a discussion of $\operatorname{Alb}(V)$ when V is defined over $\mathbb{C}$.) The Albanese variety is the maximal abelian variety into which V maps. In other words, there is a map $\pi : V \to \operatorname{Alb}(V)$ such that

for every abelian variety B and every morphism $f : V \to B$ there is a unique $f_0 : \mathrm{Alb}(V) \to B$ such that $f = f_0 \circ \pi$. If $V = A$ is an abelian variety, it is clear that $\mathrm{Alb}(A) = A$. If V is a curve, we will see in the next section that the Albanese variety is the Jacobian, hence is isomorphic to the Picard variety. In general, the connection between the Picard and Albanese varieties is as described in the following result.

Proposition A.7.3.6. *Let $\pi : V \to \mathrm{Alb}(V)$ be the universal map from V to its Albanese variety. Then the pullback map $\pi^* : \mathrm{Pic}^0(\mathrm{Alb}(V)) \to \mathrm{Pic}^0(V)$ is an isomorphism. In particular, the Albanese and Picard varieties are dual to each other.*

PROOF. See Lang [3, Section IV.4 and VI.1, Theorem 1]. □

EXERCISES

A.7.1. Let D and E be divisors on an abelian variety, and let m be a nonzero integer. Prove that $K(mD) = [m]^{-1}(K(D))$ and $K(D) \cap K(E) \subset K(E + D)$.

A.7.2. Let A be a complex abelian variety of dimension g, and let D be an ample divisor on A. Show that there exist integers $d_1, \dots, d_g$ such that $K(D) \cong (\mathbb{Z}/d_1\mathbb{Z})^2 \oplus \cdots \oplus (\mathbb{Z}/d_g\mathbb{Z})^2$ and $\ell(D) = d_1 \cdots d_g = \sqrt{\#K(D)}$. Extend this to the case where A is defined over an arbitrary field of characteristic zero.

A.7.3. This exercise provides an "analytic" proof that the group law on an abelian variety is abelian.

(a) Let G be a projective algebraic group, and for each $g \in G$ define a map $\phi(g) : G \to G$ by $\phi(g)(h) = ghg^{-1}$. The map $\phi(g)$ induces an endomorphism of the local ring $\mathcal{O}_{e,G}$, and hence an endomorphism $\phi_k(g)$ of the vector space $\mathcal{O}_{e,G}/\mathcal{M}_{e,G}^k$ for any integer k. Prove that $\phi_k(g)$ is the identity map for all k. (*Hint.* A morphism from a projective variety to an affine variety must be constant.)

(b) Use (a) to prove that $\phi(g)$ induces the identity map on $\mathcal{O}_{e,G}$. Deduce that $\phi(g)$ itself is the identity, and hence that G is commutative.

A.7.4. The purpose of this exercise is to show that a rational map from $\mathbb{P}^n$ to an abelian variety A is constant.

(a) Let G be an algebraic group, and assume that G can be embedded as a dense open subset of a smooth projective variety X. Prove that any rational map $f : G \dashrightarrow A$ must be a homomorphism followed by a translation. (*Hint.* By Corollary A.7.1.5, f is a morphism and the map $(x, y) \mapsto f(xy) - f(x) - f(y)$ extends to $X \times X \to A$. Use Lemma A.7.1.1 to finish the proof.)

(b) Now let $f : \mathbb{P}^n \dashrightarrow A$ be a rational map. Prove that f is constant. (*Hint.* Note that $\mathbb{P}^n$ contains the group $\mathbb{G}_m^n$ and the group $\mathbb{G}_a^n$.)

A.7.5. Show that the theorem of the cube (A.7.2.1) can be deduced directly from the theorem of the square (A.7.2.9) and the seesaw principle (A.7.2.3). Thus we could have used the theorem of the square as our starting point to prove the basic divisor relations on abelian varieties.

A.7.6. In this exercise we sketch the proof of the following theorem of Lang. Let X be a projective variety, let $e \in X$ be a point, and let $m : X \times X \to X$ be a morphism such that

$$m(e,x) = m(x,e) = x \qquad \text{for all } x \in X.$$

Then X is an abelian variety. To ease notation, we write $m(x,y) = xy$.
(a) Consider the map $\psi : X \times X \to X \times X$ defined by $\psi(x,y) = (xy, y)$. Show that $\psi^{-1}(e,e) = \{(e,e)\}$, and hence that ψ is onto (*Hint.* Use the dimension theorems.)
(b) Show that there exists an irreducible component Γ of the algebraic set $\{(x,y) \in X \times X \mid xy = e\}$ satisfying $p_2(\Gamma) = X$. Show that Γ also satisfies $p_1(\Gamma) = X$.
(c) Now consider the map

$$\phi : \Gamma \times X \longrightarrow X \qquad ((x',x),y) \longmapsto x'(xy).$$

Prove that $\phi((x',x),y) = y$. (*Hint.* Use the rigidity lemma.) Show also that $xx' = x'x = e$ for all $(x',x) \in \Gamma$.
(d) Use the map

$$\Gamma \times X \times X \to X, \qquad ((x',x),y,z) \mapsto x((x'y)z),$$

to show that the law is associative. Conclude that (X,m) is an algebraic group, and hence an abelian variety.

A.7.7. Let G be an algebraic group. For any (closed) subvariety V and any $g \in G$ we let $gV = \{gx \mid x \in V\}$ be the translate of V, and we define the *stabilizer of V* to be the set

$$\mathrm{Stab}_V = \{g \in G \mid gV = V\}.$$

Prove that Stab_V is a (possibly reducible) algebraic subgroup of G.

A.7.8. (Weil pairing). Let A/k be an abelian variety, let m be a positive integer, and let $\hat{a} \in \hat{A}[m]$ correspond to a divisor D in $\mathrm{Pic}^0(A)$. (If $\mathrm{char}(k) = p > 0$, we also assume that $p \nmid m$.)
(a) Show that there exist rational functions $f, g \in k(A)$ such that $mD = (f)$ and $f(mx) = g(x)^m$.
(b) Let $a \in A[m]$. Prove that the function $g(x+a)g(x)^{-1}$ is constant, that its value depends only on a and $\hat{a}$, and that the value lies in the set of mth roots of unity μ_m. We will denote this value by $e_m(a,\hat{a})$.
(c) Show that e_m is a perfect pairing $e_m : A[m] \times \hat{A}[m] \to \mu_m$.
(d) Let $c \in \mathrm{Pic}(A)$ be an ample divisor defined over k, assume that m is coprime with $\mathrm{card}(K(c))$, and let $\Phi_c : A \to \mathrm{Pic}^0(A) = \hat{A}$ be the associated polarization $\Phi_c(a) = t_a^*c - c$. Prove that the pairing

$$e_{c,m} : A[m] \times A[m] \to \mu_m, \qquad e_{c,m}(a,b) = e_m(a, \Phi_c(b)),$$

is a nondegenerate skew-symmetric pairing.
(e) Show that $k(\mu_m) \subset k(A[m])$.

A.7.9. Let A be an abelian variety, let $D \in \mathrm{Div}(A)$, and define a map $\alpha(x,y) = (x+y, x-y)$.
(a) Show that α is an isogeny from $A \times A$ to $A \times A$. What is its degree?
(b) If D is symmetric, prove that

$$\alpha^*(D \times A + A \times D) \sim 2(D \times A + A \times D).$$

(c) If D is antisymmetric, prove that $\alpha^*(D \times A + A \times D) \sim 2(D \times A)$.

A.7.10. Let A be an abelian variety, and set $\mathrm{End}_{\mathbb{Q}}(A) = \mathrm{End}(A) \otimes \mathbb{Q}$.
(a) Let $\alpha \in \mathrm{End}(A)$. Prove that α is an isogeny if and only if $\alpha \in \mathrm{End}_{\mathbb{Q}}(A)^*$.
(b) Suppose that A is simple. Prove that $\mathrm{End}_{\mathbb{Q}}(A)$ is a skew field (i.e., $\mathrm{End}_{\mathbb{Q}}(A)$ satisfies all of the axioms of a field except that its multiplication may not be commutative).
(c) Again suppose that A is simple, let $K = \mathrm{End}_{\mathbb{Q}}(A)$, and let $B = A^m$. Prove that $\mathrm{End}_{\mathbb{Q}}(B) = \mathrm{Mat}(m \times m, K)$.

A.8. Jacobians over Arbitrary Fields

In this section we develop the algebraic theory of Jacobians for smooth projective curves. The Jacobian of a curve C is an abelian variety that is naturally isomorphic to $\mathrm{Pic}^0(C)$. We sketch the construction in Section A.8.1, the main results being given in Theorems A.8.1.1 and A.8.2.1. The construction relies on some knowledge of families of varieties (Hilbert or Chow spaces), which we briefly describe in an appendix. Readers desirous of delving further into the matter should consult the survey of Milne [2] and the delightful book of Mumford [3]. The book of Serre [1] also contains the construction of Jacobians and generalized Jacobians.

A.8.1. Construction and Properties

A curve of genus 0 has $\mathrm{Pic}^0(C) = 0$, so we will concentrate on curves of positive genus. Our first theorem describes the main properties of the Jacobian of such curves.

Theorem A.8.1.1. Let C *be a smooth projective curve of genus* $g \geq 1$. *There exists an abelian variety* $\mathrm{Jac}(C)$, *called the Jacobian of* C, *and an injection* $j : C \hookrightarrow \mathrm{Jac}(C)$, *called the Jacobian embedding of* C, *with the following properties:*

(i) *Extend* j *linearly to divisors on* C. *Then* j *induces a group isomorphism between* $\mathrm{Pic}^0(C)$ *and* $\mathrm{Jac}(C)$.

(ii) *For each $r \geq 0$, define a subvariety $W_r \subset \operatorname{Jac}(C)$ by*

$$W_r = \underbrace{j(C) + \cdots + j(C)}_{r \text{ copies}}.$$

(By convention, $W_0 = \{0\}$.) Then

$$\dim(W_r) = \min(r, g) \qquad \textit{and} \qquad W_g = \operatorname{Jac}(C).$$

In particular, $\dim\bigl(\operatorname{Jac}(C)\bigr) = g$.

(iii) *Let $\Theta = W_{g-1}$. Then Θ is an irreducible ample divisor on* $\operatorname{Jac}(C)$.

Remarks. (i) It is clear that the curve C determines the pair $\bigl(\operatorname{Jac}(C), \Theta\bigr)$ up to a natural isomorphism. The converse is called Torelli's theorem: Over an algebraically closed field, the isomorphism class of the pair $\bigl(\operatorname{Jac}(C), \Theta\bigr)$ determines the isomorphism class of the curve C. See Milne [2, Theorem 12.1] for a further discussion.

(ii) Suppose that the curve C is defined over a field k. Then its Jacobian variety $\operatorname{Jac}(C)$ is also defined over k. Unfortunately, it may not be possible to define the injection $j : C \hookrightarrow \operatorname{Jac}(C)$ over k. More precisely, the map j is defined by choosing a divisor D of degree 1 and then setting

$$j : C \hookrightarrow \operatorname{Pic}^0(C) \cong \operatorname{Jac}(C), \qquad j(P) = \operatorname{Cl}\bigl((P) - D\bigr).$$

In particular, if there is a point $P_0 \in C(k)$, then we can take $D = (P_0)$ to get a map j that is defined over k. This will suffice for our proof of Faltings' theorem (Mordell conjecture), since if C has no k-rational points, then it is not difficult(!) to prove that $C(k)$ is finite. We also note that once we have identified $\operatorname{Jac}(C)$ and $\operatorname{Pic}^0(C)$, then the embedding j is unique up to translation.

(iii) There is, of course, a more intrinsic definition of the Jacobian as a variety representing the functor Pic^0. See Milne [2] for details and Exercise A.8.3 for some functoriality properties.

The fundamental tool for the algebraic construction of the Jacobian is the Riemann–Roch theorem for curves. We briefly sketch the first such construction, which is due to Weil [2, 3]. Consider the symmetric powers of the curve C,

$$\operatorname{Sym}^r C = (C \times \cdots \times C)/\mathcal{S}_r.$$

Here $\mathcal{S}_r$ denotes the symmetric group on r letters acting in the obvious way on the product of r copies of C. (See Appendix A.8.3 for more details.) We can identify $\operatorname{Sym}^r C$ with the set of effective divisors of degree r on C.

If $\operatorname{Jac}(C)$ exists, then there must be a birational morphism $\operatorname{Sym}^g C \to \operatorname{Jac}(C)$, and hence one should be able to "see" the group law on $\operatorname{Sym}^g C$. In fact, the Riemann–Roch theorem tells us that if D is a divisor of degree g, then $\ell(D) \geq 1$. Further, for "most" choices of D, there is an equality

$\ell(D) = 1$. (See Lemma A.8.2.2 below for a precise statement.) So if we fix an effective divisor $D_0 \in \operatorname{Sym}^g C$, then we can add two divisors $D_1, D_2 \in \operatorname{Sym}^g C$ by setting their sum equal to the divisor $D_3 \in \operatorname{Sym}^g C$ satisfying $D_1 + D_2 \sim D_3 + D_0$. Notice that the divisor D_3 will be uniquely determined if and only if $\ell(D_1 + D_2 - D_0) = 1$. This will define a rational map $\operatorname{Sym}^g C \times \operatorname{Sym}^g C \to \operatorname{Sym}^g C$ that satisfies the axioms of a commutative group law, except for the minor drawback that since it is only a rational map, it is not defined at all pairs of points.

Weil then proceeds to show that such a group law defined by rational maps can be transformed into an honest group law. More precisely, he shows in this situation that there exists an algebraic group G and a birational isomorphism $\rho : \operatorname{Sym}^g C \to G$ such that $\rho(x \oplus y) = \rho(x) + \rho(y)$ wherever it is defined. He then uses the valuative criterion of properness to show that G is an abelian variety, and hence that ρ must be a morphism (Corollary A.7.1.5).

We are going to present a similar construction due to Chow [1], which is perhaps less natural, but is simpler technically. The idea is to consider $\operatorname{Sym}^n C$ for some large n (it suffices to take $n \geq 2g - 1$). Then the map $\operatorname{Sym}^n C \to \operatorname{Jac}(C)$ should be a fibration, and the Riemann–Roch theorem implies that the fibers are projective spaces of dimension $n - g$. Now, all points play the same role, so we will not need any birational transformations. We obtain $\operatorname{Jac}(C)$ directly as a projective variety that parametrizes the $\mathbb{P}^{n-g}$'s lying in $\operatorname{Sym}^n C$. We now give the details of this construction, modulo some general facts about varieties parametrizing families of subvarieties that are discussed in Appendix A.8.3.

PROOF. (sketch of Theorem A.8.1.1) For simplicity we assume that C has a k-rational point $P_0 \in C(k)$. We select an integer n large enough so that for any divisor D of degree n we have $\ell(D) = n - g + 1$. (The Riemann–Roch theorem (A.4.2.3) says that any $n \geq 2g - 1$ will suffice.) Now consider the variety $\operatorname{Sym}^n C$, whose points we identify with effective divisors of degree n on C. We set $D_0 = n(P_0)$. We will use D_0 as a base point on $\operatorname{Sym}^n C$.

If $D \in \operatorname{Sym}^n C$, then the linear system $|D|$ has dimension $n - g$. Notice that the elements of $|D|$ are points of $\operatorname{Sym}^n C$, so $|D|$ is a subset of $\operatorname{Sym}^n C$. In fact, it is a subvariety. In other words, a linear system of degree n corresponds to a subvariety of $\operatorname{Sym}^n C$ that is isomorphic to $\mathbb{P}^{n-g}$. Let

$$J = \{\text{linear systems of degree } n \text{ on } C\},$$

and let π be the map

$$\pi : \operatorname{Sym}^n C \longrightarrow J, \qquad D \longmapsto |D|.$$

At this point, J is just a set, but we do know that the fibers of π are isomorphic to $\mathbb{P}^{n-g}$.

We can use our basepoint D_0 to define an addition map

$$m : J \times J \longrightarrow J, \qquad \bigl(|D_1|, |D_2|\bigr) \longmapsto |D_1 + D_2 - D_0|.$$

This construction using families of linear systems is very natural, so it is not hard to believe the following two facts (for further details, see Appendix A.8.3):

Fact 1. J is an algebraic set, and the map π is an algebraic morphism.

Fact 2. The map $m : J \times J \to J$ is an algebraic morphism.

Since $\operatorname{Sym}^n C$ is a projective variety and π is surjective, Fact 1 implies that J is a projective variety. Then the dimension theorem (A.1.3.7) and our knowledge of the fibers of π give us the dimension of J,

$$\dim(J) = \dim(\operatorname{Sym}^n C) - \dim(\mathbb{P}^{n-g}) = n - (n-g) = g.$$

Next we show that m defines a group law on J. The formulas

$$m\bigl(|D|, |D_0|\bigr) = m\bigl(|D_0|, |D|\bigr) = |D| \qquad \text{and} \qquad m\bigl(|D|, |2D_0 - D|\bigr) = |D_0|$$

show that $|D_0|$ is the identity element and that inverses exist. To check associativity, we compute

$$\begin{aligned} m\bigl(m\bigl(|D_1|, |D_2|\bigr), |D_3|\bigr) &= m\bigl(|D_1 + D_2 - D_0|, |D_3|\bigr) \\ &= |D_1 + D_2 + D_3 - 2D_0| \\ &= m\bigl(|D_1|, |D_2 + D_3 - D_0|\bigr) \\ &= m\bigl(|D_1|, m\bigl(|D_2|, |D_3|\bigr)\bigr). \end{aligned}$$

Hence m defines a group law on J, so J is an abelian variety.

We can now define a map

$$j : C \longrightarrow J, \qquad P \longmapsto \bigl|(P) + (n-1)(P_0)\bigr|,$$

and we can extend j linearly to get a map

$$j : \operatorname{Pic}^0(C) \longrightarrow J, \qquad \operatorname{class}(D) \longmapsto |D + D_0|.$$

Now the following result completes the proof of the first part of Theorem A.8.1.1.

Proposition A.8.1.2. *The map j is an isomorphism from* $\operatorname{Pic}^0(C)$ *to* J.

PROOF. Let D be a divisor of degree n. Then $j(D - D_0) = |D|$, so j is surjective. Next suppose that $j(D) = |D_0|$. This means that $|D + D_0| = |D_0|$, and hence D is a principal divisor. Therefore, j is injective. □

Next consider the set $W_r = j(C) + \cdots + j(C)$. It is the image in J of the projective variety $C \times \cdots \times C$; hence W_r is a projective variety of dimension at most r. Also, W_{r+1} is clearly equal to $W_r + j(C)$ and contains W_r, so either $W_{r+1} = W_r$ or else $\dim(W_{r+1}) = \dim(W_r) + 1$. (Note that all of

the W_r's are irreducible.) But if there is some r with $W_{r+1} = W_r$, then by induction we see that $W_s = W_r$ for all $s \geq r$. However, the surjectivity of the map $j : \mathrm{Pic}^0(C) \to J$ (A.8.1.2) tells us that the union of the W_r's fills up J, and we know from above that J has dimension g. It follows that $\dim(W_r) = r$ for all $r \leq g$, and that $\dim(W_r) = g$ for all $r \geq g$.

This completes the proof of Theorem A.8.1.1 except for the assertion that the divisor Θ is ample. We will leave this for the next section, where we will show that $K(\Theta) = \{0\}$, which implies ampleness by Theorem A.7.2.10. □

A.8.2. The Divisor Θ

The addition law on the Jacobian J is closely related to the addition of divisors on the curve C. Not surprisingly, this interplay leads to interesting divisor relations when one pulls back the theta divisor Θ to C. Similarly, one would expect an interesting divisor by pulling back the Poincaré divisor $s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta$ from $J \times J$ to $C \times C$. The next theorem describes some of these relations.

Theorem A.8.2.1. *Let C be a curve of genus g, let $P_0 \in C(k)$, and let $J = \mathrm{Jac}(C) \cong \mathrm{Pic}^0(C)$ be the Jacobian variety of C. Let $j : C \to J$ be the Jacobian embedding that sends a point P to the divisor class of $(P) - (P_0)$, and for any $c \in J$, let $j_c(P) = j(P) + c$. Further, let $\Theta = j(C) + \cdots + j(C)$ be the theta divisor on J, and let $\Theta^- = [-1]^*\Theta$.*

(i) *There is a point $\kappa \in J$ such that*

$$\Theta^- = t_\kappa^*\Theta.$$

More precisely, let K_C be a canonical divisor on C. Then $\kappa = j(K_C)$.

(ii) *With κ as in (i), for any $c \in J$ we have*

$$j_c^*\Theta^- \sim g(P_0) - c \qquad \text{and} \qquad j_c^*\Theta \sim g(P_0) - c + \kappa.$$

(iii) *Let $\Delta \subset C \times C$ be the diagonal. Then*

$$(j \times j)^*(s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta) \sim -\Delta + \big(C \times \{P_0\}\big) + \big(\{P_0\} \times C\big).$$

PROOF. Formula (i) comes from the Riemann–Roch theorem. Let D be an effective divisor of degree $g-1$, so $j(D) \in \Theta$. The Riemann–Roch theorem says that

$$\ell(K_C - D) = \deg(D) - g + 1 + \ell(D) = \ell(D) \geq 1,$$

which means that $K_C - D$ is linearly equivalent to an effective divisor E of degree $g-1$. Then $j(K_C) - j(D) = j(E) \in \Theta$, and hence $j(D) \in \Theta^- + j(K_C)$. This holds for all effective divisors D of degree $g-1$, which proves that $\Theta \subset \Theta^- + j(K_C)$. Writing $\kappa = j(K_C)$, this implies that $t_\kappa^*\Theta \subset \Theta^-$, and since they are both irreducible divisors, they must be equal. This proves (i).

In order to prove (ii), we will make use of the following lemma.

Lemma A.8.2.2. (Weil) *Let $0 \le d \le g$ be an integer. There is a nonempty open subset $U \subset \operatorname{Sym}^d C$ such that $\ell(D) = 1$ for all $D \in U$. Equivalently, the map $\operatorname{Sym}^d C \to J = \operatorname{Jac}(C)$ is injective on the set U. (By convention, we set $\operatorname{Sym}^0 C$ to be the set consisting of the divisor 0.)*

PROOF. We first observe that if D' is any effective divisor with $\ell(D') \ge 1$, then

$$\{P \in C \mid \ell(D'-P) = \ell(D') - 1\}$$

is open and nonempty. This is true because if we fix a nonzero function $f \in L(D')$, then

$$\begin{aligned} L(D'-P) = L(D') &\Longrightarrow f \in L(D'-P) \\ &\Longrightarrow \operatorname{div}(f) + D' - P \ge 0 \\ &\Longrightarrow P \in \operatorname{supp}(D') \quad \text{or} \quad f(P) = 0. \end{aligned}$$

Hence $\{P \in C \mid \ell(D'-P) = \ell(D')\}$ is contained in the union of the support of D' and the set of zeros of f, so it is finite. Now apply this with $D' = K_C - D$ for some effective divisor $D \in \operatorname{Sym}^d C$. We find that there is a nonempty open set $U \subset C$ such that

$$\ell(K_C - D) \ge 1 \Longrightarrow \ell\big(K_C - (D+P)\big) = \ell(K_C - D) - 1 \quad \text{for all } P \in U.$$

Since it is trivially true that

$$\ell(K_C - D) = 0 \Longrightarrow \ell\big(K_C - (D+P)\big) = 0 \quad \text{for all } P \in C,$$

we see by an easy induction on the degree of D that

$$\{D \in \operatorname{Sym}^d C \mid \ell(K_C - D) = g - d\}$$

is open. (Note that $\ell(K_C) = g$.) But Riemann–Roch says that this is precisely the set of D such that $\ell(D) = 1$, which completes the proof of Lemma A.8.2.2. □

We now resume the proof of Theorem A.8.2.1(ii). Let $U \subset \operatorname{Sym}^g C$ be a nonempty open set such that $\operatorname{Sym}^g C \to J$ is injective on C and such that every $D = \sum(P_i) \in U$ is a sum of distinct points P_i. Lemma A.8.2.2

tells us that such a set exists. Let $c \in -j(U)$. Then $g(P_0) - c$ is linearly equivalent to exactly one effective divisor $(P_1) + \cdots + (P_g)$, and further, the P_i's are distinct.

Now suppose that $P \in C$ is any point in the support of $j_c^*\Theta^-$. Then

$$(P) - (P_0) + c \sim -D + (g-1)(P_0)$$

for some effective divisor D of degree $g-1$, so $(P) + D \sim g(P_0) - c$. It follows that $(P) + D = (P_1) + \cdots + (P_g)$. (N.B. This is an equality of divisors, not just a linear equivalence.) Therefore, the only points that appear in $j_c^*\Theta^-$ are $P_1, \ldots, P_g$, and since they are distinct, they appear with multiplicity one. Hence $j_c^*\Theta^- = (P_1) + \cdots + (P_g) \sim g(P_0) - c$. This proves the desired result for all c in an open subset U of J.

To prove that $j_c^*\Theta^- \sim g(P_0) - c$ for all $c \in J$, we use the theorem of the square (A.7.2.9). Thus for any $a, b, c \in J$ we have

$$t_{a+b-c}^*\Theta^- \sim t_a^*\Theta^- + t_b^*\Theta^- - t_c^*\Theta^-.$$

Further, $j^* \circ t_c^* = j_c^*$, so we see that if the desired formula is true for a, b, c, then it is also true for $a + b - c$. But it is easy to see that the map from

$$U \times U \times U \longrightarrow J, \qquad (a,b,c) \longmapsto a + b - c,$$

is onto. Indeed, the map $(b, c) \mapsto b - c$ is already onto, since if $x \in J$, then $(U - x) \cap U \neq \emptyset$, and so $x = v - u$ with $u, v \in U$. This completes the proof of the first part of (ii)

To obtain the second part of (ii), we combine the first part with (i). Thus

$$j_c^*\Theta = j_c^*(t_{-\kappa}^*\Theta^-) = j_{c-\kappa}^*\Theta^- \sim g(P_0) - (c - \kappa).$$

(iii) By the seesaw principle (A.7.2.3), it suffices to prove that the two divisors are linearly equivalent on each slice $\{P\} \times C$ and $C \times \{P\}$, and by symmetry it suffices to use only the slices $\{P\} \times C$. To ease notation, we let $\delta = s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta$, and we let $i_P : C \to C \times C$ be the inclusion $i_P(Q) = (P, Q)$.

It is clear that (for $P \neq P_0$) we have

$$i_P^*\big(-\Delta + (C \times \{P_0\}) + (\{P_0\} \times C)\big) = -(P) + (P_0).$$

To compute $i_P^*(j \times j)^*\delta$, we compute each term separately. Notice that

$$p_1 \circ (j \times j) \circ i_P = \text{constant} \qquad \text{and} \qquad p_2 \circ (j \times j) \circ i_P = j,$$

so we find that

$$i_P^* \circ (j \times j)^* \circ p_1^*(\Theta) \sim 0, \qquad i_P^* \circ (j \times j)^* \circ p_2^*(\Theta) = j^*\Theta \sim g(P_0) + \kappa.$$

For the last linear equivalence we have used (ii) with $c = 0$.

Similarly,

$$(s_{12} \circ (j \times j) \circ i_P)(Q) = j(P) + j(Q) = j_{j(P)}(Q),$$

which gives

$$i_P^* \circ (j \times j)^* \circ s_{12}^*(\Theta) = j_{j(P)}^*(\Theta) \sim g(P_0) - j(P) + \kappa,$$

where again we have used (ii). Combining these calculations gives

$$\begin{aligned} i_P^* \circ (j \times j)^*(\delta) &\sim i_P^* \circ (j \times j)^* \circ s_{12}^*(\Theta) - i_P^* \circ (j \times j)^* \circ p_1^*(\Theta) \\ &\qquad - i_P^* \circ (j \times j)^* \circ p_2^*(\Theta) \\ &\sim \big(g(P_0) - j(P) + \kappa\big) - 0 - \big(g(P_0) + \kappa\big) \\ &= -j(P) \sim -(P) + (P_0). \end{aligned}$$

This completes the proof of Theorem A.8.2.1. □

We now use Theorem A.8.2.1 to show that the theta divisor on a Jacobian variety gives a principal polarization, and hence that Jacobian varieties are self dual. If we identify J with $\hat{J}$, then (A.7.3.5) says that

$$s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta \in \operatorname{Div}(J \times J)$$

defines a Poincaré divisor class. Thus Theorem A.8.2.1(iii) is really a description of the pullback of the Poincaré class from $J \times \hat{J}$ to $C \times C$. Since J is self-dual, we will generally work directly on $J \times J$ and avoid the formalism of dual abelian varieties.

Corollary A.8.2.3. *Let C be a curve of genus $g \geq 1$, let Θ be the theta divisor on its Jacobian J, and let $K(\Theta) = \{a \in J \mid t_a^*\Theta \sim \Theta\}$.*
(a) *$K(\Theta) = \{0\}$, so Θ gives a principal polarization*

$$\Phi_\Theta : J \xrightarrow{\sim} \operatorname{Pic}^0(J) \cong \hat{J}.$$

(b) *Θ is an ample divisor.*
(c) *Let $\mathcal{P}$ be a Poincaré divisor on $J \times \hat{J}$. Then*

$$(1 \times \Phi_\Theta)^*\mathcal{P} \sim s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta.$$

PROOF. (a) Let $a \in K(\Theta)$. Then $t_a^*\Theta \sim \Theta$, so two applications of Theorem A.8.2.1(ii) gives

$$g(P_0) + \kappa \sim j^*\Theta \sim j^*(t_a^*\Theta) \sim j_a^*\Theta \sim g(P_0) - a + \kappa.$$

Hence $a = 0$. This proves that $K(\Theta) = 0$, so by definition, Θ defines a principal polarization. Theorem A.7.3.1 then implies that Φ_Θ is an isomorphism.

(b) This is immediate from Theorem A.7.2.10, which says that a divisor D is ample if and only if $K(D)$ is finite.

(c) This is a restatement of Theorem A.7.3.5 and the fact that Θ defines a principal polarization. □

A.8.3. Appendix: Families of Subvarieties

We give here an introduction, via several examples, to a fundamental idea in algebraic geometry. This idea says that sets of (isomorphism classes of) varieties or maps between varieties are often themselves algebraic varieties.

One example we have already met is $\mathbb{P}^n$, which can be described as the set of lines through 0 in $\mathbb{A}^{n+1}$. Grassmann varieties (Exercise A.1.11) generalize this example. A second example is the variety $\operatorname{Sym}^n C$. This variety parametrizes effective divisors of degree n on C (i.e., unordered n-tuples of points on C). Finally, and most importantly, we have the variety of divisor classes of degree 0 on a curve C, which is precisely the Jacobian variety $\operatorname{Jac}(C)$ that we have been studying in this section. We will discuss $\operatorname{Sym}^n C$ and $\operatorname{Jac}(C)$ further below.

There is a vast literature on the general problem of *moduli spaces*, which are spaces that classify isomorphism classes of natural algebro-geometric objects. For example, the set of isomorphism classes of curves of genus g is a moduli space $\mathcal{M}_g$, and the set of isomorphism classes of principally polarized abelian varieties of dimension g is a moduli space $\mathcal{A}_g$, and both $\mathcal{M}_g$ and $\mathcal{A}_g$ have natural structures as quasi-projective varieties. We will not deal with these more difficult moduli problems, and we refer the interested reader to Mumford–Fogarty [1] for further details.

First we explain how the quotient $\operatorname{Sym}^n C = (C \times \cdots \times C)/\mathcal{S}_n$ can be given the structure of a variety. More generally, we describe the quotient of a variety by a finite group. We begin with a definition that describes what properties a quotient variety should have.

Definition. Let G be an algebraic group acting algebraically on a variety X (i.e., G is a subgroup of $\operatorname{Aut}(X)$). A *geometric quotient of X by G* is a variety Y and a morphism $\pi : X \to Y$ such that:

(1) The fibers of π are the orbits of the action of G. That is, for every $x \in X$,

$$\pi^{-1}\bigl(\pi(x)\bigr) = Gx = \{\sigma x \mid \sigma \in G\}.$$

(2) Let $f : X \to Z$ be a G-invariant morphism of varieties (i.e., $f(\sigma x) = f(x)$ for all $x \in X$ and all $\sigma \in G$). Then there is a morphism $g : Y \to Z$ such that $f = g \circ \pi$.

It is clear that if the quotient of X by G exists, then it is unique up to isomorphism. We denote the quotient, if it exists, by X/G.

The existence of such a quotient is far from automatic. One necessary condition for the existence of the quotient is that all orbits be closed. For example, the orbits of $\mathrm{GL}(n)$ acting on $\mathbb{A}^n$ are not all closed, so the quotient does not exist in this case. As the following theorem indicates, the situation is much simpler in the case of finite groups.

Theorem A.8.3.1. *The geometric quotient of a variety by a finite group exists.*

PROOF. We start with an affine variety X and a finite group $G \subset \mathrm{Aut}(X)$ and construct a morphism of affine varieties $\pi : X \to X/G$. The fundamental result from algebra that we need is a famous theorem of Hilbert.

Proposition A.8.3.2. (Hilbert) *Let A be an integral domain that is a finitely generated k-algebra. Let G be a finite group that acts on A as a k-algebra. Then the fixed subalgebra*

$$A^G = \{a \in A \,|\, \sigma(a) = a \text{ for all } \sigma \in G\}$$

is again a finitely generated k-algebra.

PROOF. Let $x_1, \ldots, x_n$ generate the k-algebra A, so $A = k[x_1, \ldots, x_n]$. Consider the polynomials

$$P_i(X) = \prod_{\sigma \in G} \bigl(X - \sigma(x_i)\bigr) = \sum_j b_{ij} X^j \in A[X].$$

The algebra $B = k[b_{11}, \ldots, b_{ng}]$ is a finitely generated k-algebra, hence is Noetherian by the Hilbert basis theorem. Further, A is integral over B by construction, and A is clearly finitely generated as a B-algebra, so A is finitely generated as a B-module.

Note that every b_{ij} is in A^G (i.e., b_{ij} is fixed by G), so A^G is a B-submodule of the finitely generated B-module A. Hence A^G is itself a finitely generated B-module, say $A^G = Bu_1 + \cdots + Bu_m$. Then $A^G = k[b_{11}, \ldots, b_{ng}, u_1, \ldots, u_m]$, which proves that A^G is a finitely generated k-algebra.

□

Now let X/k be an affine variety, and let $A = k[X] = \mathcal{O}(X)$ be its ring of regular functions. The finite group $G \subset \mathrm{Aut}(X)$ acts on A. Hilbert's theorem (A.8.3.2) tells us that the ring A^G is a finitely generated k-algebra.

Using the fact that the category of affine varieties is fully equivalent to the category of finitely generated integral k-algebras, we can find an affine variety Y/k with $k[Y] = A^G$. Then the natural inclusion $A^G \hookrightarrow A$ corresponds to a morphism $\pi : X \to Y$. We claim that Y is a geometric quotient of X by G.

By construction, π^* is the inclusion of A^G into A, so for any $\sigma \in G$ we have

$$(\pi \circ \sigma)^* = \sigma^* \circ \pi^* = \pi^*.$$

Hence $\pi \circ \sigma = \pi$, which implies that the orbits of G are contained in the fibers of π.

Next let $x, x' \in X$ have different G-orbits. Since X is affine, we can find a function $F \in k[X]$ that vanishes at x' but does not vanish at the finitely many points in the orbit Gx of x. Then the function $\prod_{\sigma \in G} F(\sigma(z))$ is in A^G, vanishes at x', and does not vanish at x. This implies that $\pi(x') \neq \pi(x)$, and hence that the fibers of π are exactly the orbits of G.

Finally, consider a G-invariant morphism $f : X \to Z$. This induces a homomorphism $f^* : k[Z] \to k[X] = A$ whose image sits in A^G, which means that f^* factors through π^*. It follows that f factors through π. This completes the proof of Theorem A.8.3.1 in the case that X is affine.

In general, if X is a quasi-projective variety, we cover X by G-invariant open affine subvarieties X_i, construct the quotients X_i/G, and glue the quotients together to obtain X/G. In order to obtain G-invariant affine open subsets, we take any hyperplane section H and note that $H' = \bigcup_{\sigma \in G} \sigma(H)$ is again a hyperplane section (the sum of very ample divisors is again very ample). Then $X \setminus H'$ is affine and G-invariant, and by varying H we can cover X with such sets. We leave the details for the reader. □

An immediate application of Theorem A.8.3.1 is that the symmetric product $\operatorname{Sym}^n V$ of any variety is again a variety. Indeed, $\operatorname{Sym}^n V$ is the quotient of the product V^n by the natural action of the symmetric group on the n coordinates. We also mention that if C is a smooth curve, then $\operatorname{Sym}^n C$ will be a smooth variety. For example, one can show that $\operatorname{Sym}^n \mathbb{A}^1 \cong \mathbb{A}^n$ and $\operatorname{Sym}^n \mathbb{P}^1 \cong \mathbb{P}^n$. However, if $\dim(V) \geq 2$, then $\operatorname{Sym}^n V$ will generally have rather nasty singularities.

As a second application, we can combine (A.8.3.1) with Poincaré's irreducibility theorem (A.5.1.7) to construct the geometric quotient of an abelian variety by an abelian subvariety. Let A be an abelian variety, and let $B \subset A$ be an abelian subvariety. Poincaré's theorem says that there is another abelian subvariety $C \subset A$ such that the map

$$s : B \times C \longrightarrow A, \qquad (b, c) \longmapsto b + c,$$

is an isogeny (i.e., s is surjective with finite kernel). Notice that A is equal to the geometric quotient of $B \times C$ by the group $\ker(s)$. We also note that $B \cap C \cong \ker(s)$ via the map $(b) \mapsto (b, -b)$, so $B \cap C$ is finite. Let Y be

the geometric quotient of C by the finite group $B \cap C$. Then the map $B \times C \to B \times Y \to Y$ factorizes through $A \cong (B \times C)/\ker(s) \to Y$, and it is easily checked that this provides a geometric quotient of A by B.

We now come to the main task of this section. Let C be a curve of genus $g \geq 1$. The points of the variety $\mathrm{Sym}^n C$ correspond to effective divisors of degree n on C, and for any $D_0 \in \mathrm{Sym}^n C$, the associated linear system

$$|D_0| = \{D \in \mathrm{Sym}^n C \mid D \sim D_0\}$$

is a subset (in fact, a subvariety) of $\mathrm{Sym}^n C$. We want to prove that the set of linear systems J can be given the structure of an algebraic variety so that the natural map

$$\mathrm{Sym}^n C \longrightarrow J, \qquad D \longmapsto |D|,$$

is a morphism. Of course, we also want to endow J with the structure of an algebraic group. Our main tool is a variant of Lemma A.8.2.2, which we state explicitly for clarity.

Lemma A.8.3.3. *Let C be a curve of genus $g \geq 1$, and let D be an effective divisor of degree $n - g$ on C. There exists a nonempty open set $U_D \subset \mathrm{Sym}^n C$ such that $\ell(D' - D) = 1$ for all $D' \in U_D$. Further, as D varies, the open sets U_D cover $\mathrm{Sym}^n C$.*

PROOF. See Milne [2, Proposition 4.2]. The proof is very similar to the proof of Lemma A.8.2.2. □

Theorem A.8.3.4. *Let C be a curve of genus $g \geq 1$, and let $n \geq 2g + 1$ be an integer. Then there exists an abelian variety J and an identification*

$$J \overset{\text{one-to-one}}{\longleftrightarrow} \{\textit{linear systems } |D| \textit{ of degree } n \textit{ on } C\}$$

such that the natural map $\pi : \mathrm{Sym}^n C \to J$, $D \mapsto |D|$, is a morphism.

PROOF. (sketch) For each effective divisor Δ of degree $n - g$ (i.e., $\Delta \in \mathrm{Sym}^{n-g}(C)$), we define two sets:

$$U_\Delta = \{D \in \mathrm{Sym}^n(C) \mid \ell(D - \Delta) = 1\},$$
$$V_\Delta = \{D \in \mathrm{Sym}^g(C) \mid D + \Delta \in U_\Delta\}.$$

Lemma A.8.3.3 says that U_Δ is open in $\mathrm{Sym}^n(C)$, and consideration of the map

$$i_\Delta : \mathrm{Sym}^g(C) \longrightarrow \mathrm{Sym}^n(C), \qquad D \longmapsto D + \Delta,$$

shows that $V_\Delta = i_\Delta^{-1}(U_\Delta)$ is open in $\mathrm{Sym}^g(C)$.

Let J be the set of linear systems of degree n on C, and consider the map

$$f_\Delta : \mathrm{Sym}^g(C) \longrightarrow J, \qquad D \longmapsto |D + \Delta|.$$

We write J_Δ for the image of f_Δ. It is clear that $\pi \circ i_\Delta = f_\Delta$ on V_Δ, and we claim that $f_\Delta : V_\Delta \to J_\Delta$ is a bijection. Indeed, if $f_\Delta(D) = f_\Delta(D')$, then $D + \Delta \sim D' + \Delta$. But by the definition of V_Δ and U_Δ, we know that $D + \Delta$ is the unique effective divisor containing Δ in its linear equivalence class. Therefore, $D + \Delta = D' + \Delta$, and hence $D = D'$.

We use the bijection $f_\Delta : V_\Delta \to J_\Delta$ to endow J_Δ with the structure of the variety V_Δ. We also note that if Δ' is another effective divisor of degree $n - g$, then f_Δ and $f_{\Delta'}$ agree on $V_\Delta \cap V_{\Delta'}$, so we can glue the algebraic structure on the J_Δ's to give all of J the structure of an algebraic variety.

The next step is to show that the map π is a morphism. We will not give the details and will just mention that this can be done by showing that π is a fibration whose fibers are isomorphic to $\mathbb{P}^{n-g}$. More precisely, we can cover $\operatorname{Sym}^n(C)$ with open sets U_i such that each U_i is isomorphic to $V_i \times \mathbb{P}^{n-g}$ for some open subset $V_i \subset J$ and such that the map $\pi : U_i \to J$ is equal to the composition

$$U_i \;\cong\; V_i \times \mathbb{P}^{n-g} \;\xrightarrow{p_1}\; V_i \;\subset J.$$

In order to describe the group law on J, we begin by observing that the map $C^n \times C^m \to C^{n+m} \to \operatorname{Sym}^{n+m} C$ is algebraic and clearly invariant by $\mathcal{S}_n \times \mathcal{S}_m$, so it induces a morphism $\operatorname{Sym}^n C \times \operatorname{Sym}^m C \to \operatorname{Sym}^{n+m} C$. The composition of this morphism with the projection $\operatorname{Sym}^{n+m} C \to J$ is invariant by linear equivalence on both factors, so it factors through an algebraic map $J \times J \to J$. (See the remark below for quotients by equivalence relations.) We will leave it to the reader to verify that this map is precisely the group law on J. This concludes our sketch of the construction of the Jacobian variety. □

Remark. More generally, one defines the *geometric quotient* of an algebraic variety X by an equivalence relation $\mathcal{R}$ to be a variety Y and a morphism $\pi : X \to Y$ satisfying:
(1) The fibers of π are the equivalence classes of $\mathcal{R}$. In other words, for each $x \in X$,

$$\pi^{-1}\bigl(\pi(x)\bigr) = \{x' \in X \mid x' \sim_{\mathcal{R}} x\}.$$

(2) Let $f : X \to Z$ be an $\mathcal{R}$-invariant morphism of varieties. Then there is a morphism $g : Y \to Z$ such that $f = g \circ \pi$.

A necessary condition for the existence of the quotient is that equivalence classes should be Zariski closed. Needless to say, it is a very difficult problem in general to give sufficient conditions for the existence. Notice that we have shown (or rather sketched) that the Jacobian variety J of C is the geometric quotient of $\operatorname{Sym}^n C$ by the linear equivalence relation on effective divisors of degree n for any fixed $n \geq 2g + 1$.

EXERCISES

A.8.1. Assume that the characteristic of k is not 2, and let $e_1, \ldots, e_{2g+1} \in k$ be distinct. Let C be the hyperelliptic curve defined by the equation $y^2 = (x-e_1)\cdots(x-e_{2g+1})$, where C includes the point ∞ at infinity.
(a) Write $P_i = (e_i, 0) \in C$. Prove that

$$\operatorname{div}(x-e_i) = 2(P_i) - 2(\infty),$$
$$\operatorname{div}(y) = (P_1) + \cdots + (P_{2g+1}) - (2g+1)(\infty).$$

(b) Let $j : C \to \operatorname{Pic}^0(C) = J$ be the embedding $j(P) = \operatorname{Cl}((P) - (\infty))$. Prove that $j(P_1), \ldots, j(P_{2g+1})$ generate the 2-torsion subgroup of J.
(c) Describe all linear relations satisfied by $j(P_1), \ldots, j(P_{2g+1})$, and use your results to prove directly that $J[2] \cong \mathbb{Z}/2^{2g}\mathbb{Z}$.

A.8.2. Let C be a smooth projective curve of genus $g \geq 1$, fix a divisor $D \in \operatorname{Div}(C)$ of degree $n \geq 1$, and use it to define a map

$$f_D : C \longrightarrow \operatorname{Pic}^0(C) = J, \qquad P \longmapsto \operatorname{Cl}(n(P) - D).$$

(a) Let $\Theta \in \operatorname{Div}(J)$ be a theta divisor on J. (Note that Θ is well-defined only up to a translation.) Prove that

$$f_D^*(\Theta + \Theta^-) \sim 2nD + n^2 K_C,$$

where $\Theta^- = [-1]^*\Theta$ and K_C is a canonical divisor on C. In particular, this divisor class is independent of the choice of Θ.
(b) Let $\mathcal{P} = s_{12}^*(\Theta) - p_1^*(\Theta) - p_2^*(\Theta)$ be the Poincaré divisor on $J \times J$ as described in (A.7.3.5). Prove that

$$(f_D \times f_D)^*(\mathcal{P}) \sim n(D \times C) + n(C \times D) - n^2\Delta.$$

(c) Take $D = K_C$, and let $\tilde{\Theta} = \Theta + \Theta^-$. Prove that

$$(f_D \times f_D)^*(g s_{12}^*\tilde{\Theta} - (g+1)p_1^*\tilde{\Theta} - (g+1)p_2^*\tilde{\Theta}) \sim -8g(g-1)^2\Delta.$$

Similarly, let $d_{12} : J \times J \to J$ be the map $d_{12}(x,y) = x - y$ and prove that

$$(g-1)(f_D \times f_D)^*(p_1^*\tilde{\Theta} + p_2^*\tilde{\Theta}) \sim g d_{12}^*\tilde{\Theta} - 8g(g-1)^2\Delta.$$

A.8.3. (Functoriality of the Jacobian) Let $\pi : C' \to C$ be a morphism between two smooth projective curves, and let $J = \operatorname{Jac}(C)$ and $J' = \operatorname{Jac}(C')$. We can use π and the identifications $J = \operatorname{Pic}^0(C)$ and $J' = \operatorname{Pic}^0(C')$ to define morphisms $\pi^* : J \to J'$ and $\pi_* : J' \to J$ as follows. The map π^* is given by pullback on divisor classes, and the map π_* is defined by the formula $\pi_*(\operatorname{Cl}(\sum n_i P_i)) = \operatorname{Cl}(\sum n_i \pi(P_i))$.
(a) Prove that π_* is a well-defined homomorphism. (*Hint.* The map π enables us to view $k(C')$ as a finite extension of $k(C)$. In particular, there is a norm map $N : k(C') \to k(C)$. Prove that if $\sum n_i P_i = \operatorname{div}(f)$, then $\sum n_i \pi(P_i) = \operatorname{div}(N(f))$.)
(b) If $\rho : C'' \to C'$ is another morphism, show that $(\pi \circ \rho)^* = \rho^* \circ \pi^*$ and $(\pi \circ \rho)_* = \pi_* \circ \rho_*$.
(c) Assume that π is nonconstant, hence surjective. Prove that π_* is surjective. Is π^* always injective? Does π^* always have a finite kernel?

A.8.4. Let k be a field with $\operatorname{char}(k) \neq 2$, and let $a, b \in k^*$. Consider the smooth projective curve C containing an affine piece U defined by the equation $y^2 = (x^2 - a^2)(x^2 - a^{-2})(x^2 - b^2)(x^2 - b^{-2})$.
(a) Show that if $(a^2 - 1)(b^2 - 1)(a^2 - b^2) \neq 0$, then U is smooth and C has genus 3.
(b) Show that the three maps

$$\begin{aligned} \phi_1 &: C \longrightarrow \mathbb{P}^2, & \phi_1(x, y) &= (x^2, y), \\ \phi_2 &: C \longrightarrow \mathbb{P}^2, & \phi_2(x, y) &= (x + x^{-1}, yx^{-2}), \\ \phi_3 &: C \longrightarrow \mathbb{P}^2, & \phi_3(x, y) &= (x - x^{-1}, yx^{-2}), \end{aligned}$$

induce morphisms of degree two from C to three elliptic curves E_1, E_2, E_3.
(c) Conclude that the Jacobian of C is isogenous to $E_1 \times E_2 \times E_3$. (*Hint.* Use the previous exercise to build a map between $\operatorname{Jac}(C)$ and $E_1 \times E_2 \times E_3$, and compute the tangent or cotangent map.)

A.8.5. Show that the Jacobian embedding $j : C \to J = \operatorname{Jac}(C)$ induces an isomorphism between regular differential 1-forms on J and regular differential 1-forms on C. (Notice that this is transparent from the analytic definition if $k = \mathbb{C}$, and it provides a bridge between the complex definition given in Section A.6 and the algebraic definition given in the present section.)

A.8.6. Let C be a smooth projective curve of genus g. The curve C is called *hyperelliptic* if there is a map $f : C \to \mathbb{P}^1$ of degree 2; it is called *trigonal* if there is a such map of degree 3; and more generally, for any $r \leq g$, the curve C is called *r-gonal* if there is a map of $\mathbb{P}^1$ of degree r. The smallest such r is sometimes called the *gonality* of the curve. Prove that the map $\operatorname{Sym}^r C \to J$ is injective if and only if the gonality of C is greater than r.

A.8.7. For any divisor class $c \in \operatorname{Pic}(C)$ of degree 1, let

$$j_c : C \longrightarrow \operatorname{Pic}^0(C) = J, \qquad P \longmapsto \operatorname{Cl}((P) - c),$$

be the associated embedding, and let $\Theta_c = j_c(C) + \cdots + j_c(C)$ be the corresponding theta divisor. Prove that there exists a c such that the divisor class of Θ_c is symmetric (i.e., $[-1]^*\Theta_c \sim \Theta_c$). How many such c's are there?

A.8.8. (a) Let $C = \mathbb{P}^1$. Prove that $\operatorname{Sym}^n C = \mathbb{P}^n$. (*Hint.* Use symmetric polynomials.)
(b) Let C be a curve of genus 1 (i.e., an elliptic curve). Prove that there is a morphism $\operatorname{Sym}^2 C \to C$ whose fibers are all isomorphic to $\mathbb{P}^1$. In other words, $\operatorname{Sym}^2 C$ is a $\mathbb{P}^1$-bundle over C.
(c) Let C be a curve of genus 2. Prove that $\operatorname{Sym}^2 C$ is isomorphic to the Jacobian of C blown up at one point.

A.8.9. (Hyperelliptic Jacobians) This exercise describes the Jacobian variety of a hyperelliptic curve. For further details, see Mumford [5, Volume 2, Chapters 1,2].

Let C be a hyperelliptic curve consisting of an affine piece given by the equation $y^2 = F(x) = (x - e_1) \cdots (x - e_{2g+1})$, together with a point ∞

at infinity. Let σ denote the involution $(x,y) \mapsto (x,-y)$ on C. Use ∞ to embed $C \hookrightarrow \operatorname{Jac}(C)$, and let Θ be the corresponding theta divisor.

(a) Show that a function whose only pole is at ∞ must have the shape $u(x) + yv(x)$ with $u, v \in k[x]$. Use this to show that if $D = \sum_{i=1}^{g} P_i$ is a divisor such that $P_i \neq \infty$ and $P_i \neq \sigma(P_j)$ for $i \neq j$, then $\ell(D) = 1$. (*Hint.* If $f \in L(D)$, then $f \prod (x - x(P_i))$ has poles only at ∞.)

(b) For each integer $r \geq 0$, define a set of effective divisors of degree r by

$$\mathcal{D}_r = \left\{ D = \sum_{i=1}^{r} P_i \;\middle|\; P_i \neq \infty \text{ and } P_i \neq \sigma(P_j) \text{ for } i \neq j \right\}.$$

Show that there is a natural identification of the set $\mathcal{D}_g$ with the set $\operatorname{Jac}(C) \smallsetminus \Theta$.

(c) Show one can give an "explicit" set of equations for the affine variety $\operatorname{Jac}(C) \smallsetminus \Theta$ as follows. For any $D = \sum (P_i) \in \mathcal{D}_g$, let $U_D(x) = \prod (x - x(P_i))$. Prove that there is a unique polynomial V_D of degree at most $g-1$ such that $y(P_i) = V_D(x(P_i))$ for all $1 \leq i \leq g$. (If P_i appears with multiplicity m in D, we impose the condition that $V_D(x) - \sqrt{F(x)}$ should vanish to order m at $x = x(P_i)$.) Further, prove that there is a unique monic polynomial W_D of degree $g+1$ such that

$$F(x) - V_D(x)^2 = U_D(x) W_D(x). \qquad (*)$$

Show that the coefficients of U, V, W, subject to the equation $(*)$, realize $\mathcal{D}_g$ as an affine subvariety in $\mathbb{A}^{3g+1}$.

(d) Recall from Exercise A.8.1 that the points $\varepsilon_i = \operatorname{Cl}((e_i, 0) - (\infty))$ generate the 2-torsion subgroup $J[2]$ of $J = \operatorname{Jac}(C)$. Prove that

$$\bigcap_{\varepsilon \in J[2]} (\Theta + \varepsilon) = \emptyset \qquad \text{and that} \qquad J = \bigcup_{\varepsilon \in J[2]} ((J \smallsetminus \Theta) + \varepsilon).$$

In other words, the translations of $J \smallsetminus \Theta$ by 2-torsion points give a covering of J by affine open sets.

(e) Prove that every divisor of degree zero is linearly equivalent to a unique divisor $D = \sum_{i=1}^{r} (P_i) - r(\infty)$ satisfying $0 \leq r \leq g$, $P_i \neq \infty$, and $P_i \neq \sigma(P_j)$ for $i \neq j$. Show that this gives a stratification of J as a disjoint union

$$J = \mathcal{D}_g \cup \cdots \cup \mathcal{D}_0.$$

Note that one can then describe the addition law as "take the sum of the two divisors and use the recipe described above to reduce it to a divisor lying in one of the $\mathcal{D}_r$'s."

A.8.10. (Generic group law on a hyperelliptic Jacobian). Show that the following procedure generically defines the group law on the Jacobian of a hyperelliptic curve. That is, it defines the group law on an open subset of $J \times J$. We retain the notation from the previous exercise. Let

$$a = \frac{g-2}{2} \text{ and } b = \frac{3g}{2} \text{ if } g \text{ is even}, \quad a = \frac{g-1}{2} \text{ and } b = \frac{3g-1}{2} \text{ if } g \text{ is odd.}$$

Let $D = \sum_{i=1}^{g}(P_i) - g(\infty)$ and $D' = \sum_{i=1}^{g}(P_i') - g(\infty)$ be divisors, and let $P_i = (x_i, y_i)$ and $P_i' = (x_i', y_i')$. Prove that there are unique polynomials A and B, with A monic of degree a and B of degree at most b, such that

$$y_i A(x_i) = B(x_i) \qquad \text{and} \qquad y_i' A(x_i') = B(x_i') \qquad \text{for all } i = 1, \dots, g.$$

Show that the function $A(x)y + B(x)$ vanishes at the $3g$ points

$$\sigma(P_1), \dots, \sigma(P_g), \sigma(P_1'), \dots, \sigma(P_g'), Q_1, \dots, Q_g,$$

where the Q_i's have the property that $D + D' \sim \sum_{i=1}^{g}(Q_i) - g(\infty)$. Using the identification from the previous problem, the divisor $\sum(Q_i)$ is thus the sum of D and D' on J. Can you give a precise description of the open subset of $J \times J$ for which this procedure is well-defined?

A.8.11. Let C be a smooth projective curve of genus g defined over the finite field $\mathbb{F}_q$ and let J be its Jacobian. Call $h = \operatorname{card}(J(\mathbb{F}_q))$ the "class number," and let δ be the smallest positive degree of a divisor rational over $\mathbb{F}_q$ (we will see that $\delta = 1$). Define

$$a_n = \operatorname{card}\{D \in \operatorname{Div}(C)_{\mathbb{F}_q} \mid D \geq 0 \text{ and } \deg(D) = n\} \quad \text{and}$$

$$Z(C/\mathbb{F}_q, T) = Z(T) = \sum_{n=0}^{\infty} a_n T^n \in \mathbb{Z}[\![T]\!].$$

(a) Start by showing that if δ does not divide n, then $a_n = 0$; whereas if δ divides n and $n \geq 2g-1$, then $a_n = h(q^{n+1-g}-1)/(q-1)$. Show also that for any divisor class c, we have $\operatorname{card}\{D \in c \mid D \geq 0\} = (q^{\ell(c)}-1)/(q-1)$.
(b) Give an expression for $Z(T)$ as a rational function of T.
(c) Let Irr_C denote the set of effective $\mathbb{F}_q$-irreducible divisors on C. Show that

$$Z(C,T) = \prod_{D \in \operatorname{Irr}_C} \left(1 - T^{\deg(D)}\right)^{-1}$$

either as a formal product or as a convergent one if $|T| < q^{-1}$.
(d) Verify that $Z(C/\mathbb{F}_{q^r}, T) = \prod_{\zeta^r=1} Z(C/\mathbb{F}_q, \zeta T)$ and use this to show that $\delta = 1$.
(e) Show that there exists a polynomial $L(C/\mathbb{F}_q, T) = L(T) \in \mathbb{Z}[T]$ such that $Z(T) = L(T)/(1-T)(1-qT)$. Show that $Z(T)$ satisfies the following functional equation:

$$Z(C/\mathbb{F}_q, T) = q^{g-1}T^{2g-2}Z(1/qT).$$

(f) Show that there exist algebraic integers $\alpha_1, \dots, \alpha_{2g}$ such that $L(T) = \prod_{i=1}^{2g}(1-\alpha_i T)$. Show that

$$\operatorname{card}(C(\mathbb{F}_{q^m})) = q^m + 1 - (\alpha_1^m + \dots + \alpha_{2g}^m) \quad \text{and} \quad h = \prod_{i=1}^{2g}(1-\alpha_i).$$

(This exercise is essentially due to F. K. Schmidt; a further property is the so-called *Riemann hypothesis* for curves over finite fields: $|\alpha_i| = \sqrt{q}$; see Hartshorne [1, Exercise V.1.10 and Appendix C].)

A.9. Schemes

This chapter is merely an introduction to the rich gallery of arithmetic schemes. We give an intrinsic meaning to the notion of good reduction and examine the minimal model of a curve and the Néron model of an abelian variety. Prerequisites for this chapter are more demanding: Most proofs in the first two sections are not that hard, but it would take too much space to fill in all the details, whereas proofs of the statements in Sections 3 and 4 are beyond the scope of this book.

A.9.1. Varieties over $\mathbb{Z}$

The idea of a scheme—a theory due entirely to one man: Grothendieck—is to abstract what we know of varieties in purely algebraic terms. A variety is covered by affine open subvarieties U_i, and to each such subvariety there corresponds a ring R_i (a finitely generated integral k-algebra). The gluing of these open subsets can be done via the sheaf $\mathcal{O}$ of regular functions, which in particular satisfies $\mathcal{O}(U_i) = R_i$. Points correspond to maximal ideals of R_i. The Zariski topology can be recovered, since a basis for the topology is given by open subsets of the type $U_{f,i} := U_i \smallsetminus \{x \mid f(x) = 0\}$; and the sheaf of regular functions is entirely characterized by $\mathcal{O}(U_{f,i}) = R_i[\frac{1}{f}]$ and by the restriction maps $R_i[\frac{1}{f}] \to R_i[\frac{1}{g}]$ for f dividing g.

We begin with the definition of an affine scheme. A first natural generalization is to drop all restrictions on the ring R: It need not be integral (it may even have nilpotent elements), nor contain a field, nor be finitely generated. A subtler shift is the passage from maximal ideals to prime ideals; one motivation for this shift is simply that the inverse image of a prime ideal is a prime ideal, whereas the same is not true for maximal ideals. For example, consider the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$; the ideal $\{0\}$ is maximal in $\mathbb{Q}$, but not in $\mathbb{Z}$.

Definition. Let R be a commutative ring. The *spectrum of* R, $\mathrm{Spec}(R)$, is a pair consisting of a topological space (by abuse of notation, also denoted by $\mathrm{Spec}(R)$) and a sheaf $\mathcal{O}$. The topological space $\mathrm{Spec}(R)$ is the set of prime ideals of R endowed with a topology whose closed sets are the sets $V(I) := \{\mathfrak{p} \in \mathrm{Spec}(R) \mid I \subset \mathfrak{p}\}$ for any ideal I of R. The sheaf $\mathcal{O}$ is characterized by $\mathcal{O}(\mathrm{Spec}(R) \smallsetminus V((f))) = R_f$ for any element $f \in R$, taken with the obvious restriction maps.

The next proposition justifies part of this construction.

Proposition A.9.1.1.
(a) *The sheaf* $\mathcal{O} = \mathcal{O}_R$ *is entirely characterized by its values on the principal open subsets* $U_f := \mathrm{Spec}(R) \smallsetminus V((f))$. *In fact, one has*

$$\mathcal{O}(U) = \varinjlim_{U_f \subset U} \mathcal{O}(U_f).$$

(b) *For* $\mathfrak{p} \in \mathrm{Spec}(R)$ *the stalk of the sheaf* $\mathcal{O}$ *at* $\mathfrak{p}$ *is (isomorphic to) the local ring* $R_{\mathfrak{p}}$.

A morphism of varieties was defined as a continuous function that sends regular functions to regular functions. We generalize this notion in the following way.

Definition.
(i) A *ringed space* is a pair $(X, \mathcal{O}_X)$ consisting of a topological space X and a sheaf of rings $\mathcal{O}_X$ on X. It is a *locally ringed space* if for all $x \in X$, the stalk $\mathcal{O}_x$ is a local ring. The sheaf $\mathcal{O}_X$ is called the *structure sheaf* of the ringed space.
(ii) A *morphism of ringed spaces* is a pair $f, f^\sharp : (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$, where $f : X \to Y$ is continuous and $f^\sharp : \mathcal{O}_Y \to f_*\mathcal{O}_X$ is a morphism of sheaves over Y, i.e., a collection of maps $f^\sharp(U) : \mathcal{O}_Y(U) \to \mathcal{O}_X(f^{-1}(U))$ such that $r_{U,V} \circ f^\sharp(U) = f^\sharp(V) \circ r_{U,V}$. It is a *morphism of locally ringed spaces* if further for all x in X, the map $f^\sharp$ induces a local ring homomorphism $f^\sharp_x : \mathcal{O}_{f(x)} \to \mathcal{O}_x$ (i.e., the inverse image of the maximal ideal is the maximal ideal).

Examples of locally ringed spaces include algebraic varieties with their sheaves of regular functions and differential (respectively analytic) varieties with their sheaves of differentiable (respectively analytic) functions.

Clearly, $(\mathrm{Spec}(R), \mathcal{O}_R)$ is a locally ringed space. These locally ringed spaces are taken as the building blocks to construct schemes.

Definition. A locally ringed space of the form $(\mathrm{Spec}(R), \mathcal{O}_R)$ is called an *affine scheme*, where R may be any ring.

Morphisms between affine schemes are described completely analogously to morphisms between affine varieties. A ring homomorphism $\phi : R \to S$ induces a morphism of locally ringed spaces $\phi^{\mathrm{sch}} = (f, f^\sharp) : (\mathrm{Spec}(S), \mathcal{O}_S) \to (\mathrm{Spec}(R), \mathcal{O}_R)$ as follows:

- If $\mathfrak{p}$ is a prime ideal of B, set $f(\mathfrak{p}) := \phi^{-1}(\mathfrak{p})$.
- If $U_g := \mathrm{Spec}(R) \smallsetminus Z(g)$, then $f^{-1}(U_g) = \mathrm{Spec}(S) \smallsetminus Z(\phi(g))$, and we set $f^\sharp(U_g) : R_g \to S_{\phi(g)}$ to be the natural map induced by ϕ on the local rings.

It is easily seen that this defines a morphism of locally ringed spaces. We formally state the converse.

Proposition A.9.1.2. *Any morphism of affine schemes* $\operatorname{Spec}(S) \to \operatorname{Spec}(R)$ *has the form* ϕ^{sch} *for some ring homomorphism* $\phi : R \to S$.

PROOF. See, for example, Hartshorne [1, Proposition II.2.3]. □

Definition. A *scheme* is a locally ringed space $(X, \mathcal{O}_X)$ that can be covered by open subsets U such that $(U, \mathcal{O}_{X_{|U}})$ is isomorphic to some affine scheme $(\operatorname{Spec}(R), \mathcal{O}_R)$. A *morphism of schemes* is a morphism of locally ringed spaces that are schemes. A scheme is called *reduced* if the rings of the structure sheaf contain no nilpotent elements, *irreducible* if the associated topological space is irreducible, and *integral* if it is both reduced and integral.

The philosophy of Grothendieck also suggests that one should always look at relative situations. This means studying *schemes over S*, or *S*-schemes, which are schemes X that come equipped with a morphism $X \to S$. In this context, if $f : X \to S$ and $g : Y \to S$ are two S-schemes, then an S-morphism is a morphism $\phi : X \to Y$ satisfying $f = g \circ \phi$. This generalizes the notion of varieties and morphisms defined over k, which corresponds to the case $S = \operatorname{Spec}(k)$. We also note that every scheme is a $\operatorname{Spec}(\mathbb{Z})$-scheme, because every ring R admits a (unique) homomorphism $\mathbb{Z} \to R$.

Examples. (a) To any affine variety X over an algebraically closed field k we can associate a k-scheme, denoted by X^{sch}, which is simply $\operatorname{Spec}(k[X])$. The closed points of X^{sch} (i.e., the maximal ideals of $k[X]$) correspond to the points of the variety X and are called *geometric points*. However, X^{sch} has many other (nonclosed) points, in fact, one for each irreducible closed subvariety of X. Of particular interest is the ideal (0), which is dense in X^{sch} and is called the *generic point of X*. Further, Proposition A.9.1.2 and Theorem A.1.2.1 say that morphisms between X and Y correspond bijectively to k-morphisms from X^{sch} to Y^{sch}, since they are both in natural bijection with the k-algebra homomorphisms from $k[Y]$ to $k[X]$.

Having turned affine varieties into schemes, it is easy to extend the construction to any quasi-projective variety X. We simply cover X by affine open sets U_i, form the affine schemes U_i^{sch}, and then glue the U_i^{sch}'s together to form the scheme X^{sch}.

(b) Of course, schemes are more general than varieties. If k is a field, the scheme $\operatorname{Spec}(k)$ has only one point. But there are other rings with only one prime ideal, for example $\mathbb{Z}/p^n\mathbb{Z}$ and $k[X]/(X^n)$ (see Exercise A.9.9). For example, the scheme $X = \operatorname{Spec}(\mathbb{Z}/p^n\mathbb{Z})$ has only one point, but it is certainly not a variety. It is irreducible, but not reduced when $n \geq 2$. Another interesting example is the spectrum of an integral local ring such as

$$\mathbb{Z}_{(p)} := \left\{ \frac{a}{b} \in \mathbb{Q} \mid b \notin p\mathbb{Z} \right\}.$$

The scheme $\operatorname{Spec}(\mathbb{Z}_{(p)})$ consists of two points, the generic point η corresponding to the ideal (0) and a unique closed point p corresponding to the

ideal $p\mathbb{Z}_{(p)}$.
(c) A scheme of fundamental importance is the affine scheme $\operatorname{Spec}(\mathbb{Z})$. It has one generic point η, corresponding to the ideal $\{0\}$, and all of its other points are closed and correspond to prime numbers,

$$\operatorname{Spec}(\mathbb{Z}) = \{(0), 2\mathbb{Z}, 3\mathbb{Z}, \dots, p\mathbb{Z}, \dots\}.$$

The structure sheaf of $\operatorname{Spec}(\mathbb{Z})$ is easy to describe:

$$\mathcal{O}(\operatorname{Spec}(\mathbb{Z}) \smallsetminus \{p_1\mathbb{Z}, \dots, p_k\mathbb{Z}\}) = \mathbb{Z}\left[\frac{1}{p_1}, \frac{1}{p_2}, \dots, \frac{1}{p_k}\right].$$

The function field of $\operatorname{Spec}(\mathbb{Z})$ (i.e., the stalk at η) is $\mathbb{Q}$. Notice that since every ring R has a canonical homomorphism $\mathbb{Z} \to R$, all schemes have a canonical morphism to $\operatorname{Spec}(\mathbb{Z})$, so every scheme is a scheme over $\operatorname{Spec}(\mathbb{Z})$.

The last example shows that $\operatorname{Spec}(\mathbb{Z})$ bears a curious resemblance to an algebraic curve. We can make this more precise by defining the dimension of a scheme.

Definition. The *dimension* of an irreducible scheme X is the maximal length n of a chain of distinct irreducible closed subsets $X_0 \subset X_1 \subset \cdots \subset X_n = X$. The *dimension* of a scheme is the maximal dimension of its irreducible components.

Examples. (a) Clearly, $\dim \operatorname{Spec}(R) = \operatorname{Krulldim}(R)$, so the dimension of a variety X is the same as the dimension of the scheme X^{sch}.
(b) The scheme of integers satisfies $\dim \operatorname{Spec}(\mathbb{Z}) = 1$, and more generally,

$$\dim \operatorname{Spec}(\mathbb{Z}[X_1, \dots, X_n]) = n + 1.$$

(c) If R is a Dedekind domain, then $\operatorname{Spec}(R)$ is irreducible, reduced, and has dimension 1.

In particular, we see that an algebraic curve and $\operatorname{Spec}(\mathbb{Z})$ are two instances of integral schemes of dimension one! Similarly, the scheme $\mathbb{A}^1_{\mathbb{Z}} := \operatorname{Spec}(\mathbb{Z}[X])$, called the "affine line over $\mathbb{Z}$", has dimension two and is analogous to $\mathbb{A}^2_k := \operatorname{Spec}(k[X, Y])$, the "affine plane over the field k."

The theory of finite coverings can thus be phrased to encompass both extensions of number fields and coverings of a curve. Field extensions $\mathbb{Q} \subset K$ and $k(C) \subset k(C')$ induce finite morphisms $\operatorname{Spec}(R_K) \to \operatorname{Spec}(\mathbb{Z})$ and $C' \to C$, and the cardinality of the fiber over a closed point is less than or equal to $[K : \mathbb{Q}]$ or $[k(C) : k(C')]$, respectively, with equality at all but finitely many points. The points where equality fails to hold are called *ramification points.* We will pursue these analogies further in the next section.

We now sketch a few salient areas where the language of schemes sheds new light on old topics or suggests new concepts and techniques.

Algebra/Geometry Schemes enable mathematicians to precisely formulate algebraic constructions in a geometric fashion. For example, the fact that $k[X, Y]$ is not principal is transparent from the fact that it corresponds to a variety of dimension two, which implies that the ideal of a closed point cannot be principal. The same argument "explains" why $\mathbb{Z}[X]$ is not principal, since it is also of dimension two. (See Exercise A.9.6 for a description of the points of $\text{Spec}(\mathbb{Z}[X])$.) Many other algebraic ideas acquire a geometric flavor. Some examples are listed in the following table.

Algebra	$\longleftrightarrow$	Geometry
ring localization	$\longleftrightarrow$	restriction to open subset
quotient ring	$\longleftrightarrow$	closed subscheme
integral closure	$\longleftrightarrow$	normalization
tensor product	$\longleftrightarrow$	geometric product

Functor of Points If X is a variety defined over a field k, a point in $X(k)$ becomes, in the language of schemes, a morphism $\text{Spec}(k) \to X^{\text{sch}}$. It is therefore natural to define a *point in X with value in S* to be a morphism $S \to X$; in other words, we define $X(S) := \text{Mor}(S, X)$. Note that S can be the spectrum of a ring, or more generally any scheme! In fancy language, the association $S \mapsto X(S)$ defines a contravariant functor from the category of schemes to the category of sets. A point x in a scheme defines a local ring $\mathcal{O}_x$, namely the stalk of the structure sheaf at x, hence a maximal ideal $\mathcal{M}_x$ and a *residue field* $k(x) := \mathcal{O}_x/\mathcal{M}_x$. In fact, a morphism $\text{Spec}(K) \to X$ is equivalent to the data of a point $x \in X$ and an injection of fields $k(x) \hookrightarrow K$. For example, one can interpret a closed k-point in a variety X as a Galois conjugacy class of points in $X(\bar{k})$.

Smoothness and Regularity A variety X is nonsingular at a point x if and only if $\dim X = \dim_k(\mathcal{M}_x/\mathcal{M}_x^2)$. Similarly, if x is a point of a scheme X, then we have a local ring $\mathcal{O}_x$ (the stalk at x of the structure sheaf) and hence a maximal ideal $\mathcal{M}_x$ and a residue field $k(x)$. We define X to be *regular* at x if $\dim X = \dim_{k(x)}(\mathcal{M}_x/\mathcal{M}_x^2)$. (The local ring is also said to be regular in that case.) Notice that the point x is not assumed to be closed, so this defines the notion of "X being nonsingular along the irreducible subvariety $Y := \overline{\{x\}}$."

Fibered Products Let $f : Y \to X$ and $g : Z \to X$ be morphisms of schemes. A *fibered product* of Y and Z over X, denoted by $Y \times_X Z$, is a scheme P with morphisms $p_1 : P \to Y$ and $p_2 : P \to Z$ such that $f \circ p_1 = g \circ p_2$ and satsifying the following universal property: For all schemes P' with morphisms $q_1 : P' \to Y$ and $q_2 : P' \to Z$ there exists a unique morphism $\phi : P' \to P$ such that $q_1 = p_1 \circ \phi$ and $q_2 = p_2 \circ \phi$. Intuitively, at least at the level of closed points, P looks like the set of

pairs (y,z) with $f(y) = g(z)$. Notice that if X, Y, and Z are varieties, then P need not be a variety; for example, it may be reducible. However, within the category of schemes, fibered product do exist.

Proposition A.9.1.3. *Let $f : Y \to X$ and $g : Z \to X$ be morphisms of schemes. Then the fibered product $Y \times_X Z$ exists and is unique up to canonical isomorphism. Further, if $X = \operatorname{Spec}(R)$, $Y = \operatorname{Spec}(A)$, and $Z = \operatorname{Spec}(B)$ are affine, then the fibered product is affine and can be described as $Y \times_X Z = \operatorname{Spec}(A \otimes_R B)$.*

PROOF. See Hartshorne [1, Chapter II, Theorem 3.3]. □

An important special case of fibered products is extension of scalars. Let X be a scheme over a ring R (i.e., X is a $\operatorname{Spec}(R)$-scheme), and let $f : R \to R'$ be a ring homomorphism. Then f induces a morphism $f^* : \operatorname{Spec}(R') \to \operatorname{Spec}(R)$, and we extend scalars on X by forming the $\operatorname{Spec}(R')$-scheme $X \times_{\operatorname{Spec}(R)} \operatorname{Spec}(R')$. To save space, people frequently say that X is an R-scheme, and write the extension as $X \times_R R'$.

Fibers of a Morphism Let $f : X \to Y$ be a morphism of schemes and let y be a not necessarily closed point of Y. The point y corresponds to a morphism $\operatorname{Spec}(k(y)) \to Y$, and we define the *fiber of f over y* to be the scheme $X_y := X \times_Y \operatorname{Spec}(k(y))$. Notice that even if y is a geometric point and if X and Y are varieties, then X_y need not be irreducible or reduced, so schemes furnish a natural language for discussing "multiple fibers." As an example, consider the hypersurface X in $\mathbb{A}^3$ defined by the equation $x^3 + ty^2 + t = 0$ and the morphism $f : X \to \mathbb{A}^1$ defined by the projection $(x, y, t) \mapsto t$. The generic fiber of f is the curve with the same equation over the field $k(t)$. For every closed point $a \in \mathbb{A}^1$ except for $a = 0$, the fiber X_a is the elliptic curve given by the equation $x^3 + ay^2 + a = 0$. However, the fiber over $a = 0$ is a triple line, $X_0 = \operatorname{Spec}(k[x, y]/(x^3))$.

Families of Schemes A *family of schemes* is just the set of fibers of a morphism of schemes $f : X \to Y$. If Y is irreducible and η is its generic point, we call $X_\eta := X \times_Y \operatorname{Spec}(k(Y))$ the *generic fiber* of the family. The fiber X_y over a closed point $y \in Y$ is called the *special fiber at y*. Notice that these definitions encompass two apparently (or at least historically) different ideas. First, if Y is an algebraic curve and X is a variety defined over an algebraically closed field k, then a family is an "algebraic deformation parametrized by a curve." The special fibers are defined over k, and the generic fiber is defined over the function field $k(Y)$. Second, if $Y = \operatorname{Spec}(\mathbb{Z})$, we get a family of schemes, where each fiber is defined over a field of different characteristic (the generic fiber being defined over $\mathbb{Q}$). Since we will be especially interested in families of schemes over curves (i.e., schemes over an algebraic curve or over $\operatorname{Spec}(\mathbb{Z})$), we state a result deeper than the ones previously quoted that describes a property of the fibers of such a family.

Proposition A.9.1.4. *(Zariski's connectedness principle) Let $f : \mathcal{X} \to S$ be an irreducible family of projective schemes over an irreducible curve S (i.e., a irreducible scheme of dimension 1). Then the generic fiber of f is irreducible. Further, every special fiber of f is connected, and all but finitely many of them are irreducible.*

PROOF. See Hartshorne [1, Chapter III, Exercise 11.4]. □

Models and Good Reduction We wish to reverse the above construction by starting with a variety X and creating a family of schemes whose generic fiber is X. Let K be either a number field or the function field $k(C)$ of a smooth projective curve, and let X be a smooth projective variety defined over K. Let $S = \operatorname{Spec}(R_K)$ if K is a number field, and let $S = C^{\text{sch}}$ if K is a function field. It is easy to see that there exists a scheme $\mathcal{X} \to R$ that is projective (by which we mean that all fibers are projective varieties) and whose generic fiber $\mathcal{X}_\eta = \mathcal{X} \times_S \operatorname{Spec}(K)$ is isomorphic to X. Indeed, fix an embedding $i : X \hookrightarrow \mathbb{P}^n_K$. We know that $\mathbb{P}^n_K$ is the generic fiber of the scheme $\mathbb{P}^n_S \to S$ (see Exercise A.9.7), so we may take $\mathcal{X}$ to be the Zariski closure of $i(X)$ inside $\mathbb{P}^n_R$. Of course, the $\mathcal{X}$ that we produce in this way may have many "bad" special fibers (e.g., reducible and/or nonreduced).

Definition. Let K and S be as above, and let X be a variety over K. A *model* for X over S is a scheme $\mathcal{X} \to S$ whose generic fiber is isomorphic to X.

For example, S is by construction a model for $\operatorname{Spec}(K)$. The scheme $\mathbb{P}^n_\mathbb{Z}$ is a model for $\mathbb{P}^n_\mathbb{Q}$ over $\operatorname{Spec}(\mathbb{Z})$ (see Exercise A.9.7). Clearly, models are not unique. One generally requires that a model have some further properties. For example, one usually insists that the morphism $\mathcal{X} \to R$ should be surjective and that each fiber should have the same dimension (the construction we sketched gives this). If X is affine, we can proceed very explicitly. Suppose that the ideal defining X in $\mathbb{A}^n$ is generated by the polynomials $P_1, \dots, P_r$. Clearing denominators, we may assume that $P_i \in \mathbb{Z}[T_1, \dots, T_n]$. Then $\operatorname{Spec}(\mathbb{Z}[T_1, \dots, T_n]/(P_1, \dots, P_r)$ gives a model for X whose special fiber at p is the scheme over $\mathbb{F}_p$ defined by the equations $\bar{P}_1 = \cdots = \bar{P}_r = 0$ in $\mathbb{A}^n_{\mathbb{F}_p}$. This illustrates that taking special fibers of a model makes precise the notion of "reducing a variety modulo p". Notice that this notion is completely intrinsic once the model is chosen (but the special fiber thus obtained may depend on the chosen model). Further, the special fiber inherits a scheme structure, so we can speak of nonreduced fibers, multiple fibers, etc.

If we are given a morphism $f : X \to Y$ defined over K and models $\mathcal{X} \to S$ and $\mathcal{Y} \to S$, it is natural to ask whether f extends to a morphism $\bar{f} : \mathcal{X} \to \mathcal{Y}$ over S. If the generic fiber is dense in $\mathcal{X}$, then there is at most one such extension, but in general we get only a rational map. For example, a linear morphism $\alpha : \mathbb{P}^n \to \mathbb{P}^n$ over $\mathbb{Q}$ extends to a morphism $\overline{\alpha} : \mathbb{P}^n_\mathbb{Z} \to \mathbb{P}^n_\mathbb{Z}$

if and only if α can be described by a matrix having integer coordinates and determinant ± 1. The following elementary result gives one reason why projective models are good.

Lemma A.9.1.5. *Let K and S be as above (i.e., S is regular of dimension 1). Let X be a variety over K, and let $\mathcal{X} \to S$ be a projective model of X. Then every K-rational point $\operatorname{Spec}(K) \to X$ of X extends to a morphism (a section) $S \to \mathcal{X}$. In other words, there is a natural bijection between $X(K)$ and $\mathcal{X}(S)$.*

PROOF. In the geometric case, the lemma follows from the fact that a rational map from a smooth curve to a projective variety is a morphism. A proof of the arithmetic case can be given along the same lines. □

It is natural, given a model $\mathcal{X} \to R$ of X/K, to say that X has good reduction at x if the fiber $\mathcal{X}_x$ is smooth. (Sometimes one adds additional requirements, for example that some endomorphism $\alpha : X \to X$ extends to $\overline{\alpha} : \mathcal{X}_x \to \mathcal{X}_x$.) To help explain the next definition, we observe that the fiber $\mathcal{X}_x$ is the same as the one obtained by first extending scalars to the local ring $\mathcal{O}_x$, next forming $\mathcal{X} \times_S S_x \to S_x = \operatorname{Spec}(\mathcal{O}_x)$, and then taking its (unique) special fiber.

Definition. A smooth projective variety X/K has *good reduction at x* if there exists a projective model of X over $\mathcal{O}_x$ whose special fiber is smooth. If such a model does not exist, we say that X has *bad reduction at x.*

Proposition A.9.1.6. *Let X be a smooth projective variety defined over K.*

(i) *The variety X has good reduction at all but finitely many points.*

(ii) *Let $T \subset S$ be the (finite) set of points where X has bad reduction. Let $S_T = \operatorname{Spec}(R_{K,T})$ in the number field case, and $S_T = C \smallsetminus T$ in the function field case. Then there exists a projective model of X over S_T all of whose fibers are smooth.*

PROOF. (sketch) (a) Let $\mathcal{X}$ be the projective model built as in the previous remarks. The smoothness of the algebraic variety X can be expressed by the nonvanishing of certain minors M_i of certain matrices with entries in K. If we let T' be the set of all points where the matrix entries have poles together with the points where the minors vanish, then $\mathcal{X}_s$ will be smooth for every point $s \in S \smallsetminus T'$.
(b) We get from (a) a smooth scheme $\mathcal{X}' \to S_{T'}$ for some finite set T' containing T. By hypothesis, for each $t \in T' \smallsetminus T$ there is a smooth scheme $\mathcal{X}_t \to S_t$. Since all these schemes have isomorphic generic fibers, we may glue them via these isomorphisms and obtain the required scheme. □

For example, $\mathbb{P}^n/\mathbb{Q}$ has good reduction everywhere, whereas the projective quartic curve $x^3y + y^3z + z^3x = 0$ has good reduction except at the point (prime) $p = 7$.

A.9.2. Analogies Between Number Fields and Function Fields

A one-dimensional affine integral regular scheme is either a smooth curve C over a field k or an open subset of the spectrum of a Dedekind ring, e.g., the ring of integers of a number field. Analogies between these two objects have fascinated many mathematicians.

Notice that in these cases the function field can be used to reconstruct the "most complete" version of the underlying scheme. Thus if we start with a field K containing an algebraically closed field k, there is a unique (up to isomorphism) smooth projective curve over k having K as its function field. Similarly, if K is a finite extension of $\mathbb{Q}$, the ring of integers R_K is the unique maximal order in K, and the associated scheme is $\operatorname{Spec}(R_K)$. We call these two situations the *geometric case* and the *arithmetic case*, respectively.

It is thus reasonable to work at the level of the field K. Valuation theory is the classical device that is used to describe the "points" of the underlying scheme.

Definition. An *absolute value* of a field K is a map $|\cdot| : K \to \mathbb{R}$ such that:

(i) $|x| \geq 0$ for all x, and $|x| = 0$ if and only if $x = 0$.

(ii) $|xy| = |x| \cdot |y|$.

(iii) $|x+y| \leq |x| + |y|$ (triangle inequality).

If further we have the stronger inequality

$$|x+y| \leq \max(|x|, |y|) \quad \text{for all } x, y \in K,$$

then the absolute value is called *ultrametric* or *nonarchimedean*. Otherwise, it is called *archimedean*. The absolute value $|x| = 1$ for all $x \neq 0$ is called the trivial absolute value.

Examples. (a) Let K be a number field. For each embedding $\sigma : K \hookrightarrow \mathbb{R}$ or $\mathbb{C}$ we get an absolute value $|x|_\sigma := |\sigma(x)|$, which is clearly archimedean. Notice that $|\cdot|_\sigma = |\cdot|_{\bar{\sigma}}$, so there are $r_1 + r_2$ of this sort. For each nonzero prime ideal $\mathfrak{p}$ of R_K we get an absolute value $|x|_\mathfrak{p} := N\mathfrak{p}^{-\operatorname{ord}_\mathfrak{p}(x)}$, which is clearly nonarchimedean.

(b) Let $K = k(T)$ with k algebraically closed (for simplicity) and T an indeterminate. For each point $a \in \mathbb{A}^1 = k$ we can similarly build a nonarchimedean absolute value by the formula $|F|_a := e^{-\operatorname{ord}_a(F)}$. There is another absolute value given by $|F|_\infty := e^{\deg(F)}$, but notice that if we introduce projective space and set $\mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\}$, then this "extra" absolute value is simply $|F|_\infty := e^{-\operatorname{ord}_\infty(F)}$. More generally, if K is the function field of a smooth projective curve C over k, then each point $a \in C$ gives a nonarchimedean absolute value $|F|_a := e^{-\operatorname{ord}_a(F)}$.

Two absolute values $|\cdot|_1$ and $|\cdot|_2$ are said to be *equivalent* if there is a real number λ such that $|\cdot|_1 = |\cdot|_2^\lambda$. It is not hard to prove for examples (a) and (b) above that every nontrivial absolute value on K is equivalent to one of the listed absolute values (provided in case (b) that it is trivial on k). We denote by M_K the set of (equivalence classes of) absolute values on K.

Theorem A.9.2.1. (Product rule) *Let K be a number field or a function field of dimension one, and for each $v \in M_K$, let $v(x) := \log|x|_v$. Then*

$$\sum_{v\in M_K} v(x) = 0 \quad \textit{for all } x \in K^*.$$

(Some might call this the "sum rule," since it is really the logarithm of the usual product rule.)

PROOF. For a number field this is the product formula (Theorem B.1.2), and for a function field it follows from the fact that a principal divisor has degree zero (see Section A.2). If the ground field is $k = \mathbb{C}$, the function field case can also be deduced from Cauchy's residue formula $\int_C dF/F = 0$. □

For any (possibly singular and/or nonprojective) curve C_0, it is possible to use the valuations of $k(C_0)$ to reconstruct a smooth projective curve that is birational to C_0. (See Hartshorne [1, Chapter 1.6], especially Theorem 6.9, for details). From this point of view, there is an important difference between the number field case and the function field case. Every curve has a natural smooth compactification, but there does not seem to be a natural compactification of Spec($\mathbb{Z}$). From the point of view of valuation theory, we should add to Spec($\mathbb{Z}$) a point ∞ corresponding to the (unique) archimedean absolute value of $\mathbb{Q}$, just as Desargues added one point to the affine line to form the projective line. (More generally, to Spec(R_K) we should add $r_1 + r_2$ points corresponding to the archimedean places of K.) Unfortunately, such an object cannot be given the structure of a scheme. Nevertheless, Arakelov has suggested a construction that enables one to translate some (but not all) theorems from the geometric case to the arithmetic case. For a brief introduction to these ideas, see Section B.10 and the references given there.

A.9.3. Minimal Model of a Curve

Let V be a variety defined over a global field K, which we assume to be either a number field or the function field of a smooth projective curve C (over a field of constants k). We let S be the scheme Spec(R_K) in the

number field case and the curve C in the function field case. The generic point of S can be identified with Spec(K). We would like to find in some sense the best possible model $\mathcal{V} \to S$ and hope that it will reflect some interesting arithmetic features of the variety V. For example, if T is the set of points of S where V has bad reduction, we will certainly require that $\mathcal{V} \times_S S_T \to S_T$ be smooth (i.e., all fibers are smooth), where recall that $S_T = S \setminus T$. That this is possible is the content of Proposition A.9.1.6; but it is not quite clear what to expect for the "bad" special fibers. A precise formulation is known only for curves and abelian varieties. In this section we describe the best models when V is a curve.

First of all, we would like the scheme $\mathcal{V}$ to be as smooth as possible; in particular, it should be regular. The other condition we want comes from the classical theory of minimal surfaces.

Definition. A projective model $\mathcal{V} \to S$ of V/K is said to be a *relatively minimal model* if it is regular and if every birational morphism from $\mathcal{V}$ to another regular model $\mathcal{V}'$ is in fact an isomorphism. The model $\mathcal{V}$ is said to be minimal if for any other regular model $\mathcal{V}'$ there is a birational morphism $\mathcal{V}' \to \mathcal{V}$.

In the geometric case, a classical result of Castelnuovo (see, for example, Hartshorne [1, Chapter V, Theorem 5.7]) states that a smooth (regular) projective surface will be a relatively minimal model of its generic fiber if and only if it contains no curves isomorphic to $\mathbb{P}^1$ having self-intersection -1. The same result is true in the arithmetic case by the work of Shafarevich [2]. The existence of a relatively minimal model for curves (in the arithmetic case) is a difficult result due to Abhyankar (desingularization) and Shafarevich (minimality). If $g \geq 1$, there is even a minimal model. The uniqueness of the minimal model is immediate from the definition.

Theorem A.9.3.1. *Let V be a curve of genus $g \geq 1$ over K. Then there exists a unique (up to isomorphism) projective minimal model $\mathcal{V} \to S$ of V.*

For example, if $\mathcal{V} \to S$ has smooth fibers, it is automatically the minimal model of its generic fiber. We thus see that a curve has good reduction if and only if the special fiber of its minimal model is smooth. If a curve has bad reduction, we can ask how bad the singularities of a singular special fiber can be.

Definition. A curve V defined over a field K (number field or 1-dimensional function field) has *semistable reduction at* $\mathfrak{p}$ if the special fiber at $\mathfrak{p}$ of the minimal model of V is reduced and has only ordinary double points as singularities.

Theorem A.9.3.2. *Let V be a smooth projective curve defined over a number field or function field K as above. There exists a finite extension L/K such that V has semistable reduction at all places L.*

PROOF. See Artin–Winter [1]. □

One can even specify a field L with the property described in Theorem A.9.3.2. For example, if K is a number field, then one may choose L to be the extension generated by torsion points of order 15 of the Jacobian of V. If K is a function field, it is enough that the ℓ-torsion points of the Jacobian of V be rational for some prime $\ell \geq 3$ coprime to char(k).

For example, the projective elliptic curve $y^2z = x^3 + tz^3$ has bad reduction at $t = 0$, but the special fiber has a singularity that is a cusp, so it is not semistable. However, over the field $k(\sqrt[6]{t})$, it is isomorphic to the curve $y^2z = x^3 + z^3$, which has good reduction at $t = 0$.

A second example is the curve $y^2z = x^3+tx^2z+t^3z^3$, which again has a cusp at $t = 0$. Over the field $k(\sqrt{t})$, it is isomorphic to $y^2z = x^3+x^2z+tz^3$, which has bad semistable reduction at $t = 0$. These two examples illustrate a general phenomenon. With a well-chosen base extension, unstable (i.e., nonsemistable) reduction becomes either good or bad semistable. For further examples, see Exercises A.9.1 to A.9.3.

A.9.4. Néron Model of an Abelian Variety

The construction of the Néron model of an abelian variety A/K follows a different path from that used in constructing good models of curves. One relaxes the properness condition and concentrates attention on the group law. In full generality, we can even dispense with the group law and simply work with morphisms, as in the following definition. We let K and S be as in the previous section (K is a number field or 1-dimensional function field, and S is a smooth scheme with generic fiber Spec(K)).

Definition. Let V/K be a variety. A scheme $\mathcal{V} \to S$ is a *Néron model of* V/K if it is smooth over S and if for every smooth scheme $\mathcal{X} \to S$ with generic fiber X/K and every morphism $f : X/K \to V/K$ it is possible to extend f to a morphism of schemes $\mathcal{X} \to \mathcal{V}$.

As usual, if a Néron model exists, it is unique up to unique isomorphism. The existence of a Néron model when V is an abelian variety is proven in a remarkable paper of Néron [1].

Theorem A.9.4.1. *Let A/K be an abelian variety. Then there exists a Néron model $\mathcal{A} \to S$ of A/K. Furthermore, $\mathcal{A}$ is a group scheme over S.*

PROOF. See Néron [1] for the original proof. Simplifications and reformulations in more modern language are given in Bosch–Lütkebohmert–Raynaud [1] and Artin [1].

□

The fact that the Néron model $\mathcal{A}$ is a group scheme is actually automatic from the definitions, since the addition map $A \times A \to A$ on the generic fiber extends to a morphism $\mathcal{A} \times_R \mathcal{A} \to \mathcal{A}$. Notice that every point $P \in A(K)$ (viewed as a morphism $\mathrm{Spec}(K) \to A$) extends to a morphism (a section) from S to $\mathcal{A}$, but that this need not be true for points in $A(L)$ for finite extensions L/K.

Example. An *abelian scheme* $\mathcal{A} \to S$ (i.e., every fiber is an abelian variety) is the Néron model of its generic fiber.

If $\mathcal{A} \to S$ is the Néron model of an abelian variety A/K, then the special fiber $\mathcal{A}_\mathfrak{p}$ over $\mathfrak{p} \in S$ will be an abelian variety if and only if it is projective, which is also equivalent to A having good reduction at $\mathfrak{p}$.

In general, the connected component of the fiber $\mathcal{A}_\mathfrak{p}$ is denoted by $\mathcal{A}_\mathfrak{p}^0$. It is an extension of an abelian variety (defined over $k_\mathfrak{p}$) by a commutative affine group. We recall that after a finite extension of the base field, any commutative affine group is isomorphic to a product of additive groups and multiplicative groups $\mathbb{G}_a^r \times \mathbb{G}_m^s$. Just as in the case of curves, taking an extension of the base field often leads to "simpler" bad special fibers; in this case it allows us to get rid of the additive groups.

Definition. An abelian variety A/K has *semistable reduction at* $\mathfrak{p}$ if the connected component of the special fiber $\mathcal{A}_\mathfrak{p}$ of the Néron model is an extension of an abelian variety by a torus T. It has *split semistable reduction* if the torus is isomorphic to $T = \mathbb{G}_m^s$ over the residue field $k_\mathfrak{p}$.

Theorem A.9.4.2. *Let A be an abelian variety defined over a number field or function field K as above. Then there exists a finite extension L/K such that A has (split) semistable reduction at all places of L.*

As in the case of curves, one can specify a field L. For number fields, one may take L to be the field generated by the torsion points of order 15 on A. For function fields, it suffices that the ℓ-torsion of A be rational for some prime $\ell \geq 3$ and coprime to $\mathrm{char}(k)$ (see Deschamps [1]).

It is natural to compare the minimal model of a curve V/K with the Néron model of its Jacobian variety $\mathrm{Jac}(V)$. For curves of genus 1, this is fairly easy.

Example A.9.4.3. Let $\mathcal{E} \to S$ be the minimal model of an elliptic curve E/K (i.e., a curve of genus 1 equipped with a rational point $P_0 \in E(K)$). Let $\mathcal{U}$ be the open subscheme of smooth points of $\mathcal{E}$. This means that $\mathcal{U}$ is obtained by discarding the multiple components and the singular points of the special fibers. Then $\mathcal{U} \to R$ is the Néron model of E/K. (See Artin [1] or Silverman [2, Chapter IV].)

More concretely, suppose $K = \mathbb{Q}$ and let

$$f(x, y, z) = -y^2z - a_1xyz - a_3yz^2 + x^3 + a_2x^2z + a_4xz^2 + a_6z^3 = 0$$

be a minimal Weierstrass equation of an elliptic curve $E/\mathbb{Q}$. We can define $\mathcal{E}$ to be the projective scheme associated to the graded ring $\mathbb{Z}[x,y,z]/(f)$ (see Exercises A.9.7 and A.9.10). The complement of the zero section is the affine scheme $\operatorname{Spec}(\mathbb{Z}[x,y]/(f(x,y,1)))$. We refer to Silverman [1, Chapter III] for the definition of the discriminant Δ and the fact that the special fiber $\tilde{E}_p := \mathcal{E} \times \mathbb{F}_p$ is smooth if and only if $\Delta \not\equiv 0 \pmod p$. If $\Delta \equiv 0 \pmod p$, then $\tilde{E}_p$ has exactly one singular point. In general, the scheme $\mathcal{E}$ will not be a minimal model for E/K, because it will not be regular. However, if $\operatorname{ord}_p(\Delta) = 0$ or 1 for every prime, then $\mathcal{E}$ is regular, hence a minimal model for E/K, and in this case the Néron model of E/K is simply $\mathcal{E}$ with the one singular point removed from each of its bad special fibers. In general, the scheme obtained by removing the singular points from the bad fibers is only the connected component of the Néron model.

For curves of higher genus, the relation between the minimal model of the curve and the Néron model of its Jacobian is considerably more complicated. We mention only a few properties, where V/K is a smooth projective curve of genus $g \geq 1$, A/K is the Jacobian variety of V, $\mathcal{V} \to S$ is the minimal model of V, and $\mathcal{A} \to S$ is the Néron model of A.

(1) The curve V/K has semistable reduction if and only if its Jacobian A/K has semistable reduction. If V/K has good reduction, then A/K also has good reduction, but the converse is not true in general.
(2) The connected component of $\mathcal{A}$ is isomorphic to $\operatorname{Pic}^0(\mathcal{V})$, the group of invertible sheaves whose restriction is of degree zero on each component of each fiber of $\mathcal{V}$. If the special fiber of $\mathcal{V}$ has components $V_1, \ldots, V_r$ that are birationally equivalent to the smooth projective curves $V_1', \ldots, V_r'$, then the abelian part of the special fiber of the Néron model of A is $\prod_i \operatorname{Jac}(V_i')$.
(3) The group of components of $\mathcal{A}$ can be easily computed from the intersection matrix of the components of the fiber of $\mathcal{V}$.

Exercises

A.9.1. Let e_i be distinct algebraic integers in a number field k and let C be the smooth hyperelliptic curve given by the affine equation $y^2 = P(x) = \prod(x - e_i)$. Let S be the set of primes dividing $2\Delta = 2(\prod_{i<j}(e_i - e_j))^2$. Prove that C has good reduction outside S. (Do not forget to check the points "at infinity.")

A.9.2. Show that the curve $y^2 = x^5 - 1$ acquires good reduction over some extension of $\mathbb{Q}$. More precisely, show that it has good reduction over the field $K = \mathbb{Q}(i, \sqrt[5]{2}, \sqrt{1 - \exp(2\pi i/5)})$. (*Hint.* Let $u = \sqrt[5]{2}$ and set $x = u^2 X$ and $y = 2Y + i$. Show that the new equation has good reduction in characteristic 2. Next let $\xi := \exp(2\pi i/5)$ and $\alpha^2 = (1-\xi)^5$, set $v = y/\alpha$ and $u = (x-1)/(1-\xi)$, and show that the new equation has the form

$v^2 = \prod(u - \varepsilon_i)$ for certain $\varepsilon_i \in \mathbb{Z}[\xi]$. Use the previous exercise to conclude that the curve has good reduction.)

A.9.3. Determine the primes of good and bad reduction over $\mathbb{Q}$ for the curves $y^2 = x^5 - x$ and $x^4 + y^4 = 1$. Do they acquire good reduction everywhere over some number field?

A.9.4. (a) Let $\mathcal{X} \to \operatorname{Spec}(\mathbb{Z})$ be a projective scheme with generic fiber $X/\mathbb{Q}$. Prove that $\mathcal{X}(\operatorname{Spec}(\mathbb{Z})) \cong X(\mathbb{Q})$.
(b) Show that this need not be true for nonprojective schemes by computing $\mathbb{G}_m(\operatorname{Spec}(\mathbb{Z}))$ and $\mathbb{G}_m(\mathbb{Q})$ and showing that they are not equal.

A.9.5. Prove the following generalization of Exercise A.1.9. Let $f : \mathcal{X} \dashrightarrow \mathcal{Y}$ be a rational map of S-schemes. Prove that there exists a scheme $\mathcal{W}$ with two morphisms $p_1 : \mathcal{W} \to \mathcal{X}$ and $p_2 : \mathcal{W} \to \mathcal{Y}$ such that p_1 is a birational map and $f \circ p_1 = p_2$. As an application prove that if $f : X \to Y$ is a morphism between two varieties defined over $\mathbb{Q}$, then there exist projective models $\mathcal{X}$ and $\mathcal{Y}$ over $\operatorname{Spec}(\mathbb{Z})$ such that f extends to a morphism $\bar{f} : \mathcal{X} \to \mathcal{Y}$.

A.9.6. (a) Show that there are three types of nonzero prime ideals in $\mathbb{Z}[X]$: (i) ideals $p\mathbb{Z}[X]$ generated by a prime number p; (ii) ideals $P(X)\mathbb{Z}[X]$ generated by a nonconstant irreducible polynomial $P(X)$; (iii) ideals $(p, P(X))$ generated by a prime number p and a nonconstant polynomial $P(X)$ whose leading coefficient is prime to p and that is irreducible when reduced modulo p. (*Hint.* Begin by considering the intersection of the ideal with $\mathbb{Z}$.)
(b) Which type(s) of ideals in (a) correspond to closed points of $\mathbb{A}^1_{\mathbb{Z}} := \operatorname{Spec}(\mathbb{Z}[X])$? More precisely, prove that an ideal of type (i) corresponds to the generic point of $\mathbb{A}^1_{\mathbb{F}_p}$, an ideal of type (ii) corresponds to a Galois conjugacy class of algebraic numbers, and an ideal of type (iii) corresponds to a Galois conjugacy class of points in $\bar{\mathbb{F}}_p$.
(c) Try to give a similar description of points in $\mathbb{A}^2_{\mathbb{Z}} := \operatorname{Spec}(\mathbb{Z}[X, Y])$.

A.9.7. (Construction of Proj of a graded ring) Let $R := \oplus_{m \geq 0} R_m$ be a graded ring. That is, R_0 is a ring, each R_m is an R_0-module, and $R_m \cdot R_n \subset R_{m+n}$. A special ideal in R is the ideal $R_+ := \oplus_{m \geq 1} R_m$. We define $\operatorname{Proj}(R)$ to be the topological space whose underlying set of points is

$$\operatorname{Proj}(R) := \{\text{homogeneous ideals } \mathfrak{p} \text{ such that } R_+ \not\subset \mathfrak{p}\}.$$

For each ideal $\mathfrak{I} \subset R$ we define a closed set

$$Z(\mathfrak{I}) := \{\mathfrak{p} \in \operatorname{Proj}(R) \mid \mathfrak{I} \subset \mathfrak{p}\}$$

in $\operatorname{Proj}(R)$, and these closed sets define the Zariski topology on $\operatorname{Proj}(R)$. For each $f \in R_+$ we also define an open set $U(f) := \operatorname{Proj}(R) \setminus Z((f))$. Then the structure sheaf $\mathcal{O} = \mathcal{O}_{\operatorname{Proj}(R)}$ on $\operatorname{Proj}(R)$ is characterized by its values

$$\mathcal{O}(U(f)) = R_{(f)} = (\text{elements of degree 0 in the local ring } R_f).$$

(a) Prove that $\operatorname{Proj}(R)$ is a scheme covered by affine open subsets $U(f)$.

(b) Let $X \subset \mathbb{P}^n$ be a projective variety, and let $S(X)$ be its homogeneous coordinate ring. Prove that $\operatorname{Proj}(S(X))$ is isomorphic to the scheme associated to the variety X.
(c) Prove that $\mathbb{P}^n_{\mathbb{Z}} := \operatorname{Proj}(\mathbb{Z}[X_0, \dots, X_n])$ is a scheme over $\operatorname{Spec}(\mathbb{Z})$ with generic fiber isomorphic (as varieties over $\mathbb{Q}$) to $\mathbb{P}^n/\mathbb{Q}$, and and special fiber over a prime p isomorphic (as varieties over $\mathbb{F}_p$) to $\mathbb{P}^n/\mathbb{F}_p$.
(d) More generally, for a ring R, define projective n-space over R to be $\mathbb{P}^n_R := \operatorname{Proj}(R[X_0, \dots, X_n])$. Show that $\mathbb{P}^n_R$ can also be described as $\mathbb{P}^n_{\mathbb{Z}} \times_{\mathbb{Z}} R$. If R is integral with fraction field K, and if η is the generic point of $\operatorname{Spec}(R)$, show that the generic fiber of $\mathbb{P}^n_R \to \operatorname{Spec}(R)$ is $\mathbb{P}^n_K$.

A.9.8. Give a description of the points of $\mathbb{P}^1_{\mathbb{Z}} := \operatorname{Proj}(\mathbb{Z}[X, Y])$ analogous to the description of the points of $\mathbb{A}^1_{\mathbb{Z}} = \operatorname{Spec}(\mathbb{Z}[X])$ given in Exercise A.9.6.

A.9.9. Let X/k be a variety. Prove that the set of tangent vectors on X is naturally isomorphic to the set of morphisms $\operatorname{Spec}(k[X]/(X^2)) \to X$. What do morphisms $\operatorname{Spec}(k[X]/(X^m))$ to X represent?

A.9.10. Let $f(x, y) := y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$ be a polynomial with $a_i \in \mathbb{Z}$ and $\Delta \neq 0$. (For the definition of the discriminant Δ we refer to Silverman [1, Chapter III].) Let $F(X, Y, Z) := ZY^2 + a_1XYZ + a_3YZ^2 - X^3 - a_2ZX^2 - a_4Z^2X - a_6Z^3$ be the corresponding homogeneous form. Let $\mathfrak{X} := \operatorname{Proj}(\mathbb{Z}[X, Y, Z]/(F))$ and $\mathfrak{U} := \operatorname{Spec}(\mathbb{Z}[x, y]/(f))$.
(a) Prove that $(X, Y, Z) = (0, 1, 0)$ defines a section $\operatorname{Spec}(\mathbb{Z}) \to \mathfrak{X}$, and that the complement of the image of this section is $\mathfrak{U}$.
(b) Prove that the special fiber $\mathfrak{X}_p$ above $p \in \operatorname{Spec}(\mathbb{Z}))$ is smooth if p does not divide Δ, and that otherwise the fiber has exactly one singular point.
(c) Prove that $\mathfrak{X}$ is regular except possibly at the singular points on its special fibers.
(d) If $\operatorname{ord}_p(\Delta) = 1$, prove that $\mathfrak{X}$ is regular even at the singular point of $\mathfrak{X}_p$.

A.9.11. (a) Show that the elliptic curve E defined by

$$F(X, Y, Z) = ZY^2 + YZ^2 - X^3 + ZX^2 = 0$$

has good reduction at all primes $p \neq 11$, and that the only singular point in characteristic 11 is $P = (8, 5, 1)$.
(b) Show that $\mathfrak{X} := \operatorname{Proj}(\mathbb{Z}[X, Y, Z]/(F))$ is a regular scheme and that $\mathfrak{X} \setminus \{P\}$ is the Néron model of E over $\mathbb{Z}$.
(c) Same questions with E' defined by

$$F(X, Y, Z) = ZY^2 + YZ^2 - X^3 + XZ^2 = 0,$$

where this time the only "bad" p is 13.

A.9.12. Let $K := \mathbb{Q}(\sqrt{29})$ and $\varepsilon := (5 + \sqrt{29})/2$.
(a) Prove that ε is a unit in $R_K = \mathbb{Z}[\varepsilon]$.
(b) Prove that the elliptic curve E defined by the affine equation

$$y^2 + xy + \varepsilon^2 y = x^3$$

has good reduction everywhere.

A.9.13. Let S be the multiplicative subset of the ring R generated by the prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_m$, and let R_S be the corresponding ring of fractions. The inclusion $R \subset R_S$ induces an inclusion of $\operatorname{Spec}(R_S)$ into $\operatorname{Spec}(R)$. Prove that $\operatorname{Spec}(R_S) = \operatorname{Spec}(R) \setminus \{\mathfrak{p}_1, \ldots, \mathfrak{p}_m\}$.

PART B

Height Functions

It is the star to every wandering bark,
Whose worth's unknown, although his height be taken.
W. Shakespeare, *Sonnet 116*

One of the fundamental tools required for the study of rational and integral points on an algebraic variety is a means of measuring the "size" of a point. A good size function will have two important attributes. First, there should be only a finite number of points of bounded size. Second, the size of a point should reflect both the arithmetic nature of the point and the geometric characteristics of the variety. The size functions that we will study in this part are called height functions. Before starting the detailed development of the theory of heights, we want to briefly amplify our description of the two properties that a good height function will possess.

For concreteness, let k be a number field and let V/k be a smooth projective variety defined over k, say with a fixed embedding $V \subset \mathbb{P}^n$. A height function corresponding to this situation will be a function

$$h : V(k) \longrightarrow [0, \infty)$$

satisfying certain properties. The finiteness property we alluded to above says that for any constant B, the set $\{P \in V(k) \mid h(P) \leq B\}$ is finite. This property lies at the heart of many of the fundamental finiteness theorems in Diophantine geometry. It is used, for example, to prove that the group of rational points on an abelian variety is finitely generated (Mordell–Weil theorem), that an affine curve of genus $g \geq 1$ has only finitely many integral points (Siegel's theorem), and that a projective curve of genus $g \geq 2$ has only finitely many rational points (Faltings' theorem). But height functions are also useful when $V(k)$ is not finite. In this case one can define the counting function

$$N\big(V(k), B\big) = \#\{P \in V(k) \mid h(P) \leq B\}.$$

Knowledge about the counting function gives arithmetic information about the variety V. For example, before Faltings' proof of the Mordell conjecture, Mumford [1] had shown that if V/k is a curve of genus $g \geq 2$, then

$N(V(k), B) \le c \log B$. This is in marked contrast to curves of genus 1, which have a counting function satisfying $c_1 B^{r/2} \le N(V(k), B) \le c_2 B^{r/2}$ for a certain integer $r \ge 0$. Vojta's proof of the Mordell conjecture is based on an extension of Mumford's argument. As a warm-up for the proof of the Mordell conjecture in Part D, we will give a proof of Mumford's theorem in this part.

The second essential property of a height function is that it should reflect the underlying geometry of the variety. More precisely, it should provide a means for translating geometric information about the variety into arithmetic information about the rational points on the variety. We will begin by defining a height function for each projective embedding of V; and then, using the relationship between projective embeddings and divisors, we will obtain an (equivalence class) of height functions for each divisor class on V. Now the geometry of V, as reflected in the structure of its divisor class group, will give corresponding information about the rational points on V. This construction, due to André Weil, is called the Height Machine. It associates to each divisor class $c \in \mathrm{Cl}(V)$ a height function $h_c : V(k) \to \mathbb{R}$, well-defined up to a bounded function on $V(k)$.

Of particular importance is the case that the variety V has some special geometric structure. For example, suppose that $V = A$ is an abelian variety. Then we can add points in A, so the height functions on A should interact in some way with the addition law. Indeed, we will prove that for an appropriate choice of $c \in \mathrm{Cl}(V)$, the corresponding height function satisfies a parallelogram law,

$$h_c(P+Q) + h_c(P-Q) = 2h_c(P) + 2h_c(Q) + O(1) \qquad \text{for all } P, Q \in A(k).$$

Here the bounded function $O(1)$ depends on the variety A, but is independent of P and Q. It follows from the parallelogram law that, up to a bounded function, the height h_c is a quadratic form on $A(k)$. In particular, $h_c(mP) = m^2 h_c(P) + O(m^2)$. These geometric properties of the height function on an abelian variety play a crucial role in the proof of the Mordell–Weil theorem.

Weil's height machine associates a height function h_c to each divisor class $c \in \mathrm{Cl}(V)$, but h_c is determined only up to a bounded function on $V(k)$. Néron and Tate showed how the group law on an abelian variety can be used to choose a particular height function $\hat{h}_c$ that has especially nice properties. For example, the parallelogram law now holds without that pesky $O(1)$,

$$\hat{h}_c(P+Q) + \hat{h}_c(P-Q) = 2\hat{h}_c(P) + 2\hat{h}_c(Q) \qquad \text{for all } P, Q \in A(k).$$

The quadratic form $\hat{h}_c$ on $A(k)$ and its associated bilinear pairing then give $A(k) \otimes \mathbb{R}$ the structure of a finite-dimensional Euclidean vector space. The group $A(k)/A(k)_{\mathrm{tors}}$ sits as a lattice inside this space, and one can then talk

about angles between points, the volume of a fundamental domain, and all of the other quantities attached to lattices in Euclidean vector spaces. All of these quantities will have a tremendous arithmetic significance because the Euclidean metric in this space is defined using the height function, and the height function itself measures the arithmetic complexity of a point. We will develop the theory of canonical heights in sections B.4 and B.5.

B.1. Absolute Values

Before we can define a size or height function on the rational points of an algebraic variety, we must first have a means of measuring the size of an algebraic number. The traditional way to describe the size of an algebraic number is through the use of absolute values. In this section we will review the theory of absolute values on number fields.

Recall that an *absolute value* on a field k is a real-valued function

$$|\cdot| : k \longrightarrow [0,\infty)$$

with the following three properties:

(1) $|x| = 0$ if and only if $x = 0$. (Nondegenerate)

(2) $|xy| = |x| \cdot |y|$. (Multiplicative)

(3) $|x+y| \le |x| + |y|$. (Triangle inequality)

The absolute value is said to be *nonarchimedean* if it satisfies

(3′) $|x+y| \le \max\{|x|, |y|\}$. (Ultrametric inequality)

We begin with the simplest number field, the field of rational numbers $\mathbb{Q}$. There is an archimedean absolute value on $\mathbb{Q}$ defined by

$$|x|_\infty = \max\{x, -x\}.$$

This is just the restriction to $\mathbb{Q}$ of the usual absolute value on $\mathbb{R}$. Further, for each prime number p there is a nonarchimedean (or p-adic) absolute value defined as follows. For any nonzero rational number $x \in \mathbb{Q}$, let $\operatorname{ord}_p(x)$ be the unique integer such that x can be written in the form

$$x = p^{\operatorname{ord}_p(x)} \cdot \frac{a}{b} \qquad \text{with } a, b \in \mathbb{Z} \text{ and } p \nmid ab.$$

(If $x = 0$, we set $\operatorname{ord}_p(x) = \infty$ by convention.) Then the p-adic absolute value of $x \in \mathbb{Q}$ is the quantity

$$|x|_p = p^{-\operatorname{ord}_p(x)}.$$

Intuitively, x is p-adically small if it is divisible by a large power of p. The homomorphism

$$\operatorname{ord}_p : \mathbb{Q}^* \longrightarrow (0,\infty)$$

is called the p-adic valuation on $\mathbb{Q}$.

Notation. The set of standard absolute values on $\mathbb{Q}$ is the set $M_{\mathbb{Q}}$ consisting of the archimedean absolute value $|\cdot|_\infty$ and the p-adic absolute values $|\cdot|_p$ for every prime p.

The set of standard absolute values on a number field k is the set M_k consisting of all absolute values on k whose restriction to $\mathbb{Q}$ is one of the standard absolute values on $\mathbb{Q}$. We write M_k^∞ for the set of archimedean absolute values in M_k, and similarly M_k^0 denotes the set of nonarchimedean absolute values on k.

To ease notation, we will frequently write the absolute value corresponding to $v \in M_k$ as $|\cdot|_v$. We also define

$$v(x) = -\log |x|_v,$$

with the convention that $v(0) = \infty$.

Let k'/k be an extension of number fields and let $v \in M_k$, $w \in M_{k'}$ be absolute values. We say that w *divides* v (or w *lies over* v) and write $w|v$ if the restriction of w to k is v. We say that v is p-adic if it lies over the p-adic absolute value of $\mathbb{Q}$.

The absolute values on $\mathbb{Q}$ satisfy the product rule

$$\prod_{v\in M_{\mathbb{Q}}} |x|_v = 1 \qquad \text{for all } x \in \mathbb{Q},\ x \neq 0.$$

This is a simple reflection of the fact that $\mathbb{Z}$ has unique factorization. In order to formulate and prove a corresponding result for number fields, we will need to assign weights to the absolute values. For any absolute value $v \in M_k$, we write k_v for the completion of the field k with respect to v. For example, let $v \in M_{\mathbb{Q}}$. Then $\mathbb{Q}_v = \mathbb{R}$ if $v = \infty$ is the archimedean absolute value on $\mathbb{Q}$, and $\mathbb{Q}_v = \mathbb{Q}_p$ if v is the p-adic absolute value. We recall the well-known formula relating the local and global degrees of an extension.

Proposition B.1.1. (Degree formula) *Let k'/k be an extension of number fields, and let $v \in M_k$ be an absolute value on k. Then*

$$\sum_{w\in M_{k'},\, w|v} [k'_w : k_v] = [k' : k].$$

PROOF. See any book on basic algebraic number theory, such as Lang [9, II, Corollary 1 to Theorem 2] or Serre [1, I, Proposition 10 and II, Théorème 1]. □

Definition. Let $v \in M_k$ be an absolute value on a number field k. The *local degree of* v is the number

$$n_v = [k_v : \mathbb{Q}_v],$$

where $\mathbb{Q}_v$ is the completion of $\mathbb{Q}$ at the restriction of v to $\mathbb{Q}$. The *normalized absolute value* associated to v is

$$\|x\|_v = |x|_v^{n_v}.$$

Proposition B.1.2. (Product formula) *Let k be a number field and let $x \in k^*$. Then*

$$\prod_{v \in M_k} \|x\|_v = 1.$$

PROOF. First we check the product formula over $\mathbb{Q}$. Write $x = \pm \prod p^{e_p}$ as a product of primes. Then

$$\prod_{v \in M_{\mathbb{Q}}} \|x\|_v = |x|_\infty \prod_p |x|_p = |x|_\infty \prod_p p^{-e_p} = 1.$$

In order to prove the product formula in general, we use the following decomposition formula. Let $x \in k$, and let $v_0 \in M_{\mathbb{Q}}$ be an absolute value on $\mathbb{Q}$. Then

$$\prod_{v \in M_k,\, v \mid v_0} \|x\|_v = \left|\mathrm{N}_{k/\mathbb{Q}}(x)\right|_{v_0}.$$

(See Lang [9, II, Corollary 2 to Theorem 2].) Using this formula, we compute

$$\prod_{v \in M_k} \|x\|_v = \prod_{v_0 \in M_{\mathbb{Q}}} \prod_{v \in M_k,\, v \mid v_0} \|x\|_v = \prod_{v_0 \in M_{\mathbb{Q}}} \left|\mathrm{N}_{k/\mathbb{Q}}(x)\right|_{v_0} = 1,$$

where the last equality follows from the product formula over $\mathbb{Q}$. □

We next give an alternative description of the absolute values on a number field k of degree $n = [k : \mathbb{Q}]$. We begin with the archimedean absolute values. It is a standard fact from field theory that k admits exactly n distinct embeddings $\sigma : k \hookrightarrow \mathbb{C}$. Each such embedding can be used to define an absolute value on k according to the rule

$$|x|_\sigma = |\sigma(x)|_\infty,$$

where $|z|_\infty$ is the usual absolute value on $\mathbb{R}$ or $\mathbb{C}$.

Recall that the embeddings $\sigma : k \hookrightarrow \mathbb{C}$ come in two flavors, the real embeddings (i.e., $\sigma(k) \subset \mathbb{R}$) and the complex embeddings (i.e., $\sigma(k) \not\subset \mathbb{R}$). The complex embeddings come in pairs that differ by complex conjugation. The usual notation is that there are r_1 real embeddings and r_2 pairs of complex embeddings, so $n = r_1 + 2r_2$. It is clear that conjugate complex embeddings give the same absolute value on k, since $|\bar{z}|_\infty = |z|_\infty$. One can show that this is the only way in which two embeddings can give the same

absolute value, and further that every archimedean absolute value arises in this way.

Next let $\mathfrak{p}$ be a prime ideal of k, say lying above the rational prime p. Also let R_k be the ring of integers of k. There is a valuation $\operatorname{ord}_\mathfrak{p}$ associated to $\mathfrak{p}$ defined by the rule that $\operatorname{ord}_\mathfrak{p}(x)$ is the exponent of $\mathfrak{p}$ in the factorization of the fractional ideal xR_k. In other words, the valuations associated to the prime ideals of k are surjective homomorphisms

$$\operatorname{ord}_\mathfrak{p} : k^* \longrightarrow \mathbb{Z} \qquad \text{characterized by} \qquad xR_k = \prod_\mathfrak{p} \mathfrak{p}^{\operatorname{ord}_\mathfrak{p}(x)}.$$

We are, of course, using the fundamental fact that R_k is a Dedekind domain, and thus that its fractional ideals have a unique factorization into a product of prime ideals.

We can use these $\mathfrak{p}$-adic valuations to define p-adic absolute values on k. Let $e_\mathfrak{p} = \operatorname{ord}_\mathfrak{p}(p)$ be the ramification index of $\mathfrak{p}$ over $\mathbb{Q}$. Then we define

$$|x|_\mathfrak{p} = p^{-\operatorname{ord}_\mathfrak{p}(x)/e_\mathfrak{p}}.$$

Notice that the $e_\mathfrak{p}$ is needed to ensure that $|p|_\mathfrak{p} = p^{-1}$. Equivalently, we can define

$$\|x\|_\mathfrak{p} = (\mathrm{N}_{k/\mathbb{Q}}\mathfrak{p})^{-\operatorname{ord}_\mathfrak{p}(x)},$$

where $\mathrm{N}_{k/\mathbb{Q}}\mathfrak{p}$ is the norm of the ideal $\mathfrak{p}$. Of course, we always understand that $\operatorname{ord}_\mathfrak{p}(0) = \infty$. We will also sometimes write

$$v_\mathfrak{p}(x) = -\log |x|_\mathfrak{p}$$

when we are feeling in an additive, rather than a multiplicative, mood.

We summarize the above discussion in the following proposition.

Proposition B.1.3. *Let $k/\mathbb{Q}$ be a number field of degree $n = [k : \mathbb{Q}]$.*
(a) *Let*

$$\rho_1, \ldots, \rho_{r_1} : k \hookrightarrow \mathbb{R} \qquad \textit{and} \qquad \tau_1, \bar{\tau}_1, \ldots, \tau_{r_2}, \bar{\tau}_{r_2} : k \hookrightarrow \mathbb{C}$$

be the real and complex embeddings of k, respectively. Then there is a bijection

$$\{\rho_1, \rho_2, \ldots, \rho_{r_1}, \tau_1, \tau_2, \ldots, \tau_{r_2}\} \xleftrightarrow{\sim} M_k^\infty, \qquad \sigma \longmapsto |\cdot|_\sigma,$$

where $|x|_\sigma = |\sigma(x)|_\infty$ is the absolute value described above.
(b) *Let p be a prime, and let $pR_k = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$ be the factorization of p in the ring of integers of k. Then there is a bijection*

$$\{\mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_r\} \xleftrightarrow{\sim} \{p\textit{-adic absolute values on } k\}, \qquad \mathfrak{p} \longmapsto |\cdot|_\mathfrak{p},$$

where $|x|_\mathfrak{p} = p^{-\operatorname{ord}_\mathfrak{p}(x)/e_\mathfrak{p}}$ *is the absolute value described above.*

Notice that one consequence of Proposition B.1.3 is that a number field k has one absolute value for each prime ideal, one absolute value for each real embedding, and one absolute value for each pair of complex embeddings. The nonarchimedean absolute values M_k^0 are those corresponding to the prime ideals. The ring of integers of k can be characterized using absolute values as

$$R_k = \{x \in k \mid |x|_v \le 1 \text{ for all } v \in M_k^0\}.$$

More generally, if $S \subset M_k$ is any set of absolute values containing the archimedean absolute values M_k^∞, then the *ring of S-integers of k* is defined to be

$$R_S = \{x \in k \mid |x|_v \le 1 \text{ for all } v \in M_k,\ v \notin S\}.$$

Thus with this terminology, R_k is the ring of M_k^∞-integers of k.

B.2. Heights on Projective Space

There is a natural way to measure the size of a rational point $P \in \mathbb{P}^n(\mathbb{Q})$. Such a point can be written (almost uniquely) in the form

$$P = (x_0, x_1, \dots, x_n) \quad \text{with } x_0, x_1, \dots, x_n \in \mathbb{Z} \text{ and } \gcd(x_0, x_1, \dots, x_n) = 1.$$

We define the *height of P* to be the quantity

$$H(P) = \max\{|x_0|, |x_1|, \dots, |x_n|\}.$$

It is clear that for any B, the set

$$\{P \in \mathbb{P}^n(\mathbb{Q}) \mid H(P) \le B\}$$

is finite, since there are only finitely many integers $x \in \mathbb{Z}$ satisfying $|x| \le B$.

This notion of height can be generalized to number fields in the following way.

Definition. Let k be a number field, and let $P = (x_0, x_1, \dots, x_n) \in \mathbb{P}^n(k)$ be a point whose homogeneous coordinates are chosen in k. The *height of P (relative to k)* is the quantity

$$H_k(P) = \prod_{v \in M_k} \max\{\|x_0\|_v, \|x_1\|_v, \dots, \|x_n\|_v\}.$$

We also define

$$h_k(P) = \log H_k(P) = \sum_{v \in M_k} -n_v \min\{v(x_0), v(x_1), \dots, v(x_n)\}.$$

In order to distinguish between them, we call H_k the *multiplicative height* and h_k the *logarithmic height.*

The product formula ensures that the height $H_k(P)$ is well-defined, independent of the choice of homogeneous coordinates for P. We verify this and describe the dependence on k in the following lemma.

Lemma B.2.1. *Let k be a number field and let $P \in \mathbb{P}^n(k)$ be a point.*
(a) *The height $H_k(P)$ is independent of the choice of homogeneous coordinates for P.*
(b) *$H_k(P) \geq 1$ for all $P \in \mathbb{P}^n(k)$.*
(c) *Let k' be a finite extension of k. Then*

$$H_{k'}(P) = H_k(P)^{[k':k]}.$$

PROOF. (a) Write $P = (x_0, \dots, x_n)$. Any other choice of coordinates for P has the form $(cx_0, \dots, cx_n)$ with $c \in k^*$. Using the product formula (B.1.2), we find that

$$\begin{aligned}\prod_{v \in M_k} \max\{\|cx_0\|_v, \dots, \|cx_n\|_v\} \\ &= \Bigl(\prod_{v \in M_k} \|c\|_v\Bigr)\Bigl(\prod_{v \in M_k} \max\{\|x_0\|_v, \dots, \|x_n\|_v\}\Bigr) \\ &= \prod_{v \in M_k} \max\{\|x_0\|_v, \dots, \|x_n\|_v\}.\end{aligned}$$

(b) We can take homogeneous coordinates for P such that some coordinate is equal to 1. Then it is clear from the definition that $H_k(P) \geq 1$.
(c) For this part we use the degree formula to compute

$$\begin{aligned} H_{k'}(P) &= \prod_{w \in M_{k'}} \max\{\|x_0\|_w, \dots, \|x_n\|_w\} \\ &= \prod_{v \in M_k} \prod_{w \in M_{k'},\, w|v} \max\{\|x_0\|_w, \dots, \|x_n\|_w\} \\ &= \prod_{v \in M_k} \prod_{w \in M_{k'},\, w|v} \max\{|x_0|_v^{n_w}, \dots, |x_n|_v^{n_w}\}. \end{aligned}$$

Now $n_w = [k'_w : \mathbb{Q}_w] = [k'_w : k_v]n_v$, so we get

$$\begin{aligned} &= \prod_{v \in M_k} \prod_{w \in M_{k'},\, w|v} \max\{\|x_0\|_v, \dots, \|x_n\|_v\}^{[k'_w:k_v]} \\ &= \prod_{v \in M_k} \max\{\|x_0\|_v, \dots, \|x_n\|_v\}^{[k':k]} \quad \text{from (B.1.1)} \\ &= H_k(P)^{[k':k]}. \end{aligned}$$

□

The transformation formula (B.2.1(c)) allows us to define a height function that is independent of the field.

Definition. The *absolute (multiplicative) height* on $\mathbb{P}^n$ is the function

$$H : \mathbb{P}^n(\bar{\mathbb{Q}}) \longrightarrow [1, \infty), \qquad H(P) = H_k(P)^{1/[k:\mathbb{Q}]},$$

where k is any field with $P \in \mathbb{P}^n(k)$. The *absolute (logarithmic) height* on $\mathbb{P}^n$ is

$$h : \mathbb{P}^n(\bar{\mathbb{Q}}) \longrightarrow [0, \infty), \qquad h(P) = \log H(P) = \frac{1}{[k:\mathbb{Q}]} h_k(P).$$

Note that (B.2.1(c)) ensures that $H(P)$ is well-defined independent of the choice of the field k.

We also define the height of an element $\alpha \in k$ to be the height of the corresponding projective point $(\alpha, 1) \in \mathbb{P}^1(k)$. Thus

$$H_k(\alpha) = \prod_{v \in M_k} \max\{\|\alpha\|_v, 1\},$$

and similarly for $h_k(\alpha)$, $H(\alpha)$, and $h(\alpha)$.

Proposition B.2.2. *The action of the Galois group on $\mathbb{P}^n(\bar{\mathbb{Q}})$ leaves the height invariant. In other words, let $P \in \mathbb{P}^n(\bar{\mathbb{Q}})$ and let $\sigma \in G_{\mathbb{Q}}$. Then $H(\sigma(P)) = H(P)$.*

PROOF. Let $k/\mathbb{Q}$ be a number field with $P \in \mathbb{P}^n(k)$. The automorphism σ of $\bar{\mathbb{Q}}$ defines an isomorphism $\sigma : k \xrightarrow{\sim} \sigma(k)$, and it likewise identifies the sets of absolute values on k and $\sigma(k)$. More precisely,

$$\sigma : M_k \xrightarrow{\sim} M_{\sigma(k)}, \qquad v \longmapsto \sigma(v),$$

where for $x \in k$ and $v \in M_k$, the absolute value $\sigma(v) \in M_{\sigma(k)}$ is defined by $|\sigma(x)|_{\sigma(v)} = |x|_v$. It is also clear that σ induces an isomorphism on the completions, $k_v \cong \sigma(k)_{\sigma(v)}$, so $n_v = n_{\sigma(v)}$. This allows us to compute

$$\begin{aligned} H_{\sigma(k)}(\sigma(P)) &= \prod_{w \in M_{\sigma(k)}} \max\{\|\sigma(x_i)\|_w\} = \prod_{w \in M_{\sigma(k)}} \max\{|\sigma(x_i)|_w\}^{n_w} \\ &= \prod_{v \in M_k} \max\{|\sigma(x_i)|_{\sigma(v)}\}^{n_{\sigma(v)}} = \prod_{v \in M_k} \max\{|x_i|_v\}^{n_v} \\ &= \prod_{v \in M_k} \max\{\|x_i\|_v\} = H_k(P). \end{aligned}$$

We also have $[k:\mathbb{Q}] = [\sigma(k):\mathbb{Q}]$, so taking $[k:\mathbb{Q}]^{\text{th}}$ roots gives the desired result. □

Recall that the field of definition of a point $P = (x_0, \ldots, x_n) \in \mathbb{P}^n(\bar{\mathbb{Q}})$ is the field

$$\mathbb{Q}(P) = \mathbb{Q}(x_0/x_j, x_1/x_j, \ldots, x_n/x_j) \qquad \text{for any } j \text{ with } x_j \neq 0.$$

The following finiteness theorem is of fundamental importance for the application of height functions in Diophantine geometry.

Theorem B.2.3. *For any numbers $B, D \geq 0$, the set*

$$\{P \in \mathbb{P}^n(\bar{\mathbb{Q}}) \mid H(P) \leq B \text{ and } [\mathbb{Q}(P) : \mathbb{Q}] \leq D\}$$

is finite. In particular, for any fixed number field k, the set

$$\{P \in \mathbb{P}^n(k) \mid H_k(P) \leq B\}$$

is finite.

PROOF. Choose homogeneous coordinates for $P = (x_0, \ldots, x_n)$ such that some coordinate equals 1. Then for any absolute value v and any index i we have

$$\max\{\|x_0\|_v, \ldots, \|x_n\|_v\} \geq \max\{\|x_i\|_v, 1\}.$$

Multiplying over all v and taking an appropriate root, we see that

$$H(P) \geq H(x_i) \qquad \text{for all } 0 \leq i \leq n.$$

Further, it is clear that $\mathbb{Q}(P) \supset \mathbb{Q}(x_i)$. Hence it suffices to prove that for each $1 \leq d \leq D$, the set

$$\{x \in \bar{\mathbb{Q}} \mid H(x) \leq B \text{ and } [\mathbb{Q}(x) : \mathbb{Q}] = d\}$$

is finite.

Let $x \in \bar{\mathbb{Q}}$ have degree d and let $k = \mathbb{Q}(x)$. We write $x_1, \ldots, x_d$ for the conjugates of x over $\mathbb{Q}$, and we let

$$F_x(T) = \prod_{j=1}^{d}(T - x_j) = \sum_{r=0}^{d}(-1)^r s_r(x) T^{d-r}$$

be the minimal polynomial of x over $\mathbb{Q}$. For any absolute value $v \in M_k$, we can estimate the size of the symmetric polynomial $s_r(x)$ by

$$\begin{aligned} \left|s_r(x)\right|_v &= \left| \sum_{1 \leq i_1 < \cdots < i_r \leq d} x_{i_1} \cdots x_{i_r} \right|_v \\ &\leq c(v,r,d) \max_{1 \leq i_1 < \cdots < i_r \leq d} |x_{i_1} \cdots x_{i_r}|_v \quad \text{(triangle inequality)} \\ &\leq c(v,r,d) \max_{1 \leq i \leq d} |x_i|_v^r. \end{aligned}$$

Here $c(v,r,d) = \binom{d}{r} \leq 2^d$ if v is archimedean, and we can take $c(v,r,d) = 1$ if v is nonarchimedean.

It follows that

$$\max\{\left|s_0(x)\right|_v, \ldots, \left|s_d(x)\right|_v\} \leq c(v,d) \prod_{i=1}^{d} \max\{|x_i|_v, 1\}^d,$$

where $c(v,d) = 2^d$ if v is archimedean, and $c(v,d) = 1$ otherwise. Now we multiply this inequality over all $v \in M_k$ and take the $[k:\mathbb{Q}]^{\text{th}}$ root to obtain the estimate

$$H\big(s_0(x), \ldots, s_d(x)\big) \le 2^d \prod_{i=1}^{d} H(x_i)^d.$$

But the x_i's are conjugates, so (B.2.2) tells us that all of the $H(x_i)$'s are equal. Hence

$$H\big(s_0(x), \ldots, s_d(x)\big) \le 2^d H(x)^{d^2}.$$

Now suppose that x is in the set

$$\{x \in \bar{\mathbb{Q}} \mid H(x) \le B \text{ and } [\mathbb{Q}(x):\mathbb{Q}] = d\}.$$

Then we have just proven that x is the root of a polynomial $F_x(T) \in \mathbb{Q}[T]$ whose coefficients $s_0, \ldots, s_d$ satisfy $H(s_0, \ldots, s_d) \le 2^d B^{d^2}$. But we saw earlier that $\mathbb{P}^d(\mathbb{Q})$ has only finitely many points of bounded height, so there are only finitely many possibilities for the polynomial $F_x(T)$, and hence only finitely many possibilities for x. This completes the proof of Theorem B.2.3. □

An immediate corollary of the finiteness property in Theorem B.2.3 is the following important result due to Kronecker. We will later prove a generalization, see (B.4.3(a)).

Corollary B.2.3.1. (Kronecker's theorem) *Let k be a number field, and let $P = (x_0, \ldots, x_n) \in \mathbb{P}^n(k)$. Fix any i with $x_i \ne 0$. Then $H(P) = 1$ if and only if the ratio x_j/x_i is a root of unity or zero for every $0 \le j \le n$.*

PROOF. Without loss of generality, we may divide the coordinates of P by x_i and then reorder them, so we may assume that $P = (1, x_1, x_2, \ldots, x_n)$. First suppose that every x_j is a root of unity. Then $|x_j|_v = 1$ for every absolute value on k, and hence $H(P) = 1$.

Next suppose that $H(P) = 1$. For each $r = 1, 2, \ldots$, let $P^r = (x_0^r, \ldots, x_n^r)$. It is clear from the definition of the height that $H(P^r) = H(P)^r$, so $H(P^r) = 1$ for every $r \ge 1$. But $P^r \in \mathbb{P}^n(k)$, so Theorem B.2.3 tells us that the sequence $P, P^2, P^3, \ldots$ contains only finitely many distinct points. Choose integers $s > r \ge 1$ such that $P^s = P^r$. This implies that $x_j^s = x_j^r$ for each $1 \le j \le n$ (since we have dehomogenized with $x_0 = 1$). Therefore, each x_j is a root of unity or is zero. □

The next two results give our first examples of the interplay between geometry and arithmetic. The proof of Proposition B.2.4 is elementary, while the proof of Theorem B.2.5 uses the Nullstellensatz and the triangle inequality to translate the geometric assertion that a map is a morphism into an arithmetic relationship between height functions.

Proposition B.2.4. *Let $S_{n,m}$ be the Segre embedding described in Example A.1.2.6(b),*

$$S_{n,m} : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^N, \qquad (x, y) \longmapsto (x_0y_0, x_0y_1, \ldots, x_iy_j, \ldots, x_ny_m).$$

Let H_n, H_m, and H_N be hyperplanes in $\mathbb{P}^n$, $\mathbb{P}^m$, and $\mathbb{P}^N$, respectively.

(a) $$S^*_{n,m}(H_N) \sim H_n \times \mathbb{P}^m + \mathbb{P}^n \times H_m \in \operatorname{Div}(\mathbb{P}^n \times \mathbb{P}^m).$$

(b) $h\big(S_{n,m}(x, y)\big) = h(x) + h(y)$ *for all $x \in \mathbb{P}^n(\bar{\mathbb{Q}})$ and $y \in \mathbb{P}^m(\bar{\mathbb{Q}})$.*

(c) *Let $\Phi_d : \mathbb{P}^n \to \mathbb{P}^N$ be the d-uple embedding described in Example A.1.2.6(a). Then*

$$h\big(\Phi_d(x)\big) = dh(x) \quad \textit{for all } x \in \mathbb{P}^n(\bar{\mathbb{Q}}).$$

PROOF. (a) Let $(z_0, \ldots, z_N)$ be homogeneous coordinates on $\mathbb{P}^N$, and fix hyperplanes $H_N = \{z_0 = 0\}$, $H_n = \{x_0 = 0\}$, and $H_m = \{y_0 = 0\}$. Then

$$\begin{aligned} S^*_{n,m}H_N &= S^*_{n,m}\{z \in \mathbb{P}^N \mid z_0 = 0\} \\ &= \big\{(x, y) \in \mathbb{P}^n \times \mathbb{P}^m \mid x_0y_0 = 0\big\} = H_n \times \mathbb{P}^m + \mathbb{P}^n \times H_m. \end{aligned}$$

(b) Let $x \in \mathbb{P}^n(k)$ and $y \in \mathbb{P}^m(k)$ for some number field k, and let $z = S_{m,n}(x, y)$. Then for any absolute value $v \in M_k$ we have

$$\max_{0 \le \ell \le N} |z_\ell|_v = \max_{\substack{0 \le i \le n \\ 0 \le j \le m}} |x_iy_j|_v = \left(\max_{0 \le i \le n} |x_i|_v\right)\left(\max_{0 \le j \le m} |y_j|_v\right).$$

Now raise to the $n_v/[k : \mathbb{Q}]$ power, multiply over all $v \in M_k$, and take logarithms to obtain the desired result.

(c) The d-uple embedding is defined by $\Phi(x) = (M_0(x), \ldots, M_N(x))$, where the $M_i(x)$ are all monomials of degree d in $n + 1$ variables. It is clear that $\big|M_i(x)\big|_v \le \max_i |x_i|_v^d$, and since the particular monomials $x_0^d, \ldots, x_n^d$ appear in the list, we find that

$$\max_{0 \le j \le N} \big|M_j(x)\big|_v = \max_{0 \le i \le n} |x_i|_v^d.$$

Now raise to the $n_v/[k : \mathbb{Q}]$ power, multiply over all $v \in M_k$, and take logarithms to finish the proof. □

Theorem B.2.5. *Let $\phi : \mathbb{P}^n \to \mathbb{P}^m$ be a rational map of degree d defined over $\bar{\mathbb{Q}}$, so ϕ is given by an $(n+1)$-tuple $\phi = (f_0, \ldots, f_m)$ of homogeneous polynomials of degree d. Let $Z \subset \mathbb{P}^n$ be the subset of common zeros of the f_i's. Notice that ϕ is defined on $\mathbb{P}^n \smallsetminus Z$.*

(a) *We have*

$$h(\phi(P)) \le dh(P) + O(1) \qquad \text{for all } P \in \mathbb{P}^n(\bar{\mathbb{Q}}) \smallsetminus Z.$$

(b) *Let X be a closed subvariety of $\mathbb{P}^n$ with the property that $X \cap Z = \emptyset$. (Thus ϕ defines a morphism $X \to \mathbb{P}^m$.) Then*

$$h(\phi(P)) = dh(P) + O(1) \qquad \text{for all } P \in X(\bar{\mathbb{Q}}).$$

We also record a special case that will be needed in the next section.

Corollary B.2.6. *Let $A : \mathbb{P}^n \to \mathbb{P}^m$ be a linear map defined over $\bar{\mathbb{Q}}$. In other words, A is given by $m+1$ linear forms $(L_0, \ldots, L_m)$. Let $Z \subset \mathbb{P}^n$ be the linear subspace where $L_0, \ldots, L_m$ simultaneously vanish, and let $X \subset \mathbb{P}^n$ be a closed subvariety with $X \cap Z = \emptyset$. Then*

$$h(A(P)) = h(P) + O(1) \qquad \text{for all } P \in X(\bar{\mathbb{Q}}).$$

Remark B.2.7. (i) The $O(1)$'s in Theorem B.2.5 depend on the map ϕ, but are independent of the point P.
(ii) It is possible to give an explicit formula for the $O(1)$'s in terms of the coefficients of ϕ and the equations defining X. More precisely, it is quite easy to get an explicit upper bound $h(\phi(P)) \le dh(P) + c_1(\phi)$ for (a) using only the triangle inequality; see Exercise B.1. The corresponding lower bound $h(\phi(P)) \ge dh(P) - c_2(\phi)$ is more difficult, even in the case that $Z = \emptyset$. See, for example, the effective Nullstellensatz proven by Masser and Wüstholz [1].
(2) It is not true in general that $h(\phi(P)) \ge dh(P) + O(1)$ for all points $P \in \mathbb{P}^n(\bar{\mathbb{Q}}) \smallsetminus Z$. See Exercise B.2 for an example.

PROOF (of Theorem B.2.5 and Corollary B.2.6). Fix a field of definition k for ϕ, so

$$\phi = (f_0, f_1, \ldots, f_m) \qquad \text{with } f_0, \ldots, f_m \in k[X_0, \ldots, X_n]_d.$$

(That is, the f_i's are homogeneous polynomials of degree d.) We write f_i explicitly as

$$f_i(X) = \sum_{|e|=d} a_{i,e} X^e,$$

where $e = (e_0, \ldots, e_n)$ is a multi-index, $|e| = e_0 + \cdots + e_n$, and $X^e = X_0^{e_0} X_1^{e_1} \cdots X_n^{e_n}$. Notice that this sum has $\binom{n+d}{n}$ terms, which is the number of monomials of degree d in $n+1$ variables.

For any point $P = (x_0, \ldots, x_n)$ with $x_j \in k$ and any absolute value $v \in M_k$, we will write $|P|_v = \max\{|x_j|_v\}$. Similarly, for any polynomial $f = \sum a_e X^e \in k[X]$ we will let $|f|_v = \max\{|a_e|\}$. We also set the convenient notation

$$\varepsilon_v(r) = \begin{cases} r & \text{if } v \text{ is archimedean,} \\ 1 & \text{if } v \text{ is nonarchimedean.} \end{cases}$$

With this notation, the triangle inequality can be written uniformly as

$$|a_1 + a_2 + \cdots + a_r|_v \le \varepsilon_v(r) \max\{|a_1|_v, \ldots, |a_r|_v\}.$$

Now consider any point $P \in \mathbb{P}^n(\bar{\mathbb{Q}})$. Extending k if necessary, we may assume that $P \in \mathbb{P}^n(k)$ and write $P = (x_0, \ldots, x_n)$ with $x_i \in k$. Then for any $v \in M_k$ and any i we have

$$\begin{aligned} |f_i(P)|_v &= \left| \sum_{|e|=d} a_{i,e} x^e \right|_v \\ &\le \varepsilon_v \binom{n+d}{n} \left(\max_e |a_{i,e}|_v\right) \left(\max_e |x_0^{e_0} \cdots x_n^{e_n}|_v\right) \\ &\le \varepsilon_v \binom{n+d}{n} |f_i|_v \max_{e,j} |x_j|_v^{e_0 + \cdots + e_n} \\ &= \varepsilon_v \binom{n+d}{n} |f_i|_v |P|_v^d. \end{aligned}$$

Now take the maximum over $0 \le i \le m$, raise to the $n_v/[k:\mathbb{Q}]$ power, and multiply over all $v \in M_k$. This gives

$$H(\phi(P)) \le \binom{n+d}{n} H(\phi) H(P)^d,$$

where we are writing $H(\phi)$ for the quantity

$$H(\phi) = \prod_{v \in M_k} \max\{|f_0|_v, \ldots, |f_m|_v\}^{n_v/[k:\mathbb{Q}]}.$$

(That is, $H(\phi)$ is the height of the point $(a_{i,e})$ whose coordinates are the coefficients of all of the f_i's.) We have also made use of the identity

$$\prod_{v \in M_k} \varepsilon_v(r)_v^n = \prod_{v \in M_k^\infty} r_v^n = r^{[k:\mathbb{Q}]},$$

which follows from the degree formula (B.1.1). Taking logarithms gives

$$h(\phi(P)) \le dh(P) + h(\phi) + \log \binom{n+d}{n},$$

which completes the proof of (a).

(b) In order to get a complementary inequality, we need to use the fact that we are choosing points $P \in X$ and that ϕ is a morphism on X. Let $p_1, \ldots, p_r$ be homogeneous polynomials generating the ideal of X. Then we know that $p_1, \ldots, p_r, f_0, \ldots, f_m$ have no common zeros in $\mathbb{P}^n$. The Nullstellensatz (Theorem A.1.1.2) tells us that the ideal they generate has a radical equal to the ideal generated by $X_0, X_1, \ldots, X_n$. This means that we can find polynomials g_{ij}, q_{ij} (which we may assume to be homogeneous) and an exponent $t \geq d$ such that

$$g_{0j}f_0 + \cdots + g_{mj}f_m + q_{1,j}p_1 + \cdots + q_{r,j}p_r = X_j^t \qquad \text{for } 0 \leq j \leq n.$$

Notice that the g_{ij}'s are homogeneous of degree $t-d$, since the f_i's are homogeneous of degree d. Extending k if necessary, we may also assume that the g_{ij}'s and q_{ij}'s have coefficients in k. Now let $P = (x_0, \ldots, x_n) \in X(k)$. The assumption that $P \in X$ implies that $p_i(P) = 0$ for all i, so when we evaluate the above formula at P we obtain

$$g_{0j}(P)f_0(P) + \cdots + g_{mj}(P)f_m(P) = x_j^t, \qquad 0 \leq j \leq n.$$

Hence

$$\begin{aligned}
|P|_v^t &= \max_j |x_j^t|_v \\
&= \max_j \bigl|g_{0j}(x)f_0(x) + g_{1j}(x)f_1(x) + \cdots + g_{mj}(x)f_m(x)\bigr|_v \\
&\leq \varepsilon_v(m+1)\bigl(\max_{i,j} |g_{ij}(x)|_v\bigr)\bigl(\max_i |f_i(x)|_v\bigr) \\
&\leq \varepsilon_v(m+1)\left[\varepsilon_v\binom{t-d+n}{n}\bigl(\max_{i,j}|g_{ij}|_v\bigr)|P|_v^{t-d}\right] \cdot \bigl(\max_i |f_i(x)|_v\bigr).
\end{aligned}$$

Now raise to the $n_v/[k:\mathbb{Q}]$ power and multiply over all $v \in M_k$. This yields

$$H(P)^t \leq cH(P)^{t-d}H\bigl(\phi(P)\bigr),$$

where c is a certain constant depending on the f_i's, the g_{ij}'s, and t, but independent of P. In other words, c depends only on ϕ and X, so taking logarithms gives the desired inequality

$$dh(P) \leq h\bigl(\phi(P)\bigr) + O(1).$$

This completes the proof of Theorem B.2.5. □

B.3. Heights on Varieties

Let V be a projective variety defined over $\bar{\mathbb{Q}}$. If V is embedded in some $\mathbb{P}^n$, or more generally if we are given a morphism $\phi : V \to \mathbb{P}^n$, then we can define a height function on V.

Definition. Let $\phi : V \to \mathbb{P}^n$ be a morphism. The (*absolute logarithmic*) *height on V relative to ϕ* is the function

$$h_\phi : V(\bar{\mathbb{Q}}) \longrightarrow [0,\infty), \qquad h_\phi(P) = h\big(\phi(P)\big),$$

where $h : \mathbb{P}^n(\bar{\mathbb{Q}}) \to [0,\infty)$ is the height function on projective space defined in the previous section.

We have seen (B.2.5) that if $\phi : \mathbb{P}^n \to \mathbb{P}^m$ is any morphism of degree d, then $h_\phi(P) = dh(P) + O(1)$. The fact that ϕ has degree d is equivalent to the assertion that $\phi^*H' \sim dH$, where H and H' are hyperplanes in $\mathbb{P}^n$ and $\mathbb{P}^m$, respectively. The following strengthening of (B.2.5) will be one of the crucial ingredients in the construction of the height machine later in this section.

Theorem B.3.1. *Let V be a projective variety defined over $\bar{\mathbb{Q}}$, let $\phi : V \to \mathbb{P}^n$ and $\psi : V \to \mathbb{P}^m$ be morphisms, and let H and H' be hyperplanes in $\mathbb{P}^n$ and $\mathbb{P}^m$, respectively. Suppose that ϕ^*H and ψ^*H' are linearly equivalent (i.e., ϕ and ψ are associated to the same complete linear system). Then*

$$h_\phi(P) = h_\psi(P) + O(1) \qquad \text{for all } P \in V(\bar{\mathbb{Q}}).$$

Here the $O(1)$ constant will depend on V, ϕ, and ψ, but it is independent of P.

PROOF. Let $D \in \operatorname{Div}(V)$ be any positive divisor in the linear equivalence class of ϕ^*H and ψ^*H'. The morphisms ϕ and ψ are determined respectively by certain subspaces V and V' in the vector space $L(D)$ and choices of bases for V and V'. (See Section A.3.1.) In other words, if we choose a basis $h_0, \dots, h_N$ for $L(D)$, then there are linear combinations

$$f_i = \sum_{j=0}^{N} a_{ij}h_j, \quad 0 \le i \le n,$$
$$g_i = \sum_{j=1}^{N} b_{ij}h_j, \quad 0 \le i \le m,$$

such that ϕ and ψ are given by

$$\phi = (f_0, \dots, f_n) \qquad \text{and} \qquad \psi = (g_0, \dots, g_m).$$

Here the a_{ij}'s and b_{ij}'s are constants.

Let $\lambda = (h_0, \ldots, h_N) : V \to \mathbb{P}^N$ be the morphism corresponding to the complete linear system determined by D. Let A be the linear map $A : \mathbb{P}^N \to \mathbb{P}^n$ defined by the matrix (a_{ij}), and similarly let $B : \mathbb{P}^N \to \mathbb{P}^m$ be the linear map defined by (b_{ij}). Then we have commutative diagrams

$$\begin{array}{ccc} V & \xrightarrow{\lambda} & \mathbb{P}^N \\ & \phi \searrow & \downarrow A \\ & & \mathbb{P}^n \end{array} \qquad\qquad \begin{array}{ccc} V & \xrightarrow{\lambda} & \mathbb{P}^N \\ & \psi \searrow & \downarrow B \\ & & \mathbb{P}^m \end{array}$$

The vertical maps A and B are not morphisms on all of $\mathbb{P}^N$, but the fact that ϕ and ψ are morphisms associated to the linear system $L(D)$ implies that A is defined at every point of the image $\lambda(V(\bar{\mathbb{Q}}))$, and similarly for B. Hence we can apply Corollary B.2.6 to conclude that

$$h(A(Q)) = h(Q) + O(1) \quad \text{and} \quad h(B(Q)) = h(Q) + O(1) \qquad \text{for all } Q \in \lambda(V(\bar{\mathbb{Q}})).$$

Writing $Q = \lambda(P)$ with $P \in V(\bar{\mathbb{Q}})$ and using the commutative diagrams gives the desired result:

$$\begin{aligned} h(\phi(P)) &= h(A(\lambda(P)) = h(\lambda(P)) + O(1) \\ &= h(B(\lambda(P))) + O(1) = h(\psi(P)) + O(1). \end{aligned}$$

□

We are now ready to give Weil's construction that associates a height function to every divisor. This theorem may be viewed as a machine that converts geometric statements described in terms of divisor class relations into arithmetic statements described by relations between height functions.

Theorem B.3.2. (Weil's Height Machine) *Let k be a number field. For every smooth projective variety V/k there exists a map*

$$h_V : \operatorname{Div}(V) \longrightarrow \{\text{functions } V(\bar{k}) \to \mathbb{R}\}$$

with the following properties:

(a) (Normalization) *Let $H \subset \mathbb{P}^n$ be a hyperplane, and let $h(P)$ be the absolute logarithmic height on $\mathbb{P}^n$ defined in Section B.2. Then*

$$h_{\mathbb{P}^n,H}(P) = h(P) + O(1) \qquad \text{for all } P \in \mathbb{P}^n(\bar{k}).$$

(b) (Functoriality) *Let $\phi : V \to W$ be a morphism and let $D \in \operatorname{Div}(W)$. Then*

$$h_{V,\phi^*D}(P) = h_{W,D}\big(\phi(P)\big) + O(1) \qquad \text{for all } P \in V(\bar{k}).$$

(c) (Additivity) *Let $D, E \in \operatorname{Div}(V)$. Then*

$$h_{V,D+E}(P) = h_{V,D}(P) + h_{V,E}(P) + O(1) \qquad \text{for all } P \in V(\bar{k}).$$

(d) (Linear Equivalence) *Let $D, E \in \operatorname{Div}(V)$ with D linearly equivalent to E. Then*

$$h_{V,D}(P) = h_{V,E}(P) + O(1) \qquad \text{for all } P \in V(\bar{k}).$$

(e) (Positivity) *Let $D \in \operatorname{Div}(V)$ be an effective divisor, and let B be the base locus of the linear system $|D|$. Then*

$$h_{V,D}(P) \geq O(1) \qquad \text{for all } P \in (V \smallsetminus B)(\bar{k}).$$

(f) (Algebraic Equivalence) *Let $D, E \in \operatorname{Div}(V)$ with D ample and E algebraically equivalent to 0. Then*

$$\lim_{\substack{P \in V(\bar{k}) \\ h_{V,D}(P) \to \infty}} \frac{h_{V,E}(P)}{h_{V,D}(P)} = 0.$$

(See also Theorem B.5.9 for a stronger statement.)

(g) (Finiteness) *Let $D \in \operatorname{Div}(V)$ be ample. Then for every finite extension k'/k and every constant B, the set*

$$\{P \in V(k') \mid h_{V,D}(P) \leq B\}$$

is finite.

(h) (Uniqueness) *The height functions $h_{V,D}$ are determined, up to $O(1)$, by normalization (a), functoriality (b) just for embeddings $\phi : V \hookrightarrow \mathbb{P}^n$, and additivity (c).*

Remarks B.3.2.1. (i) If the variety V is not smooth, Weil's Height Machine (B.3.2) is still valid, provided that one works entirely with Cartier divisors, rather than with Weil divisors. The proof of Theorem B.3.2 for singular (projective) varieties goes through verbatim using the theory of Cartier divisors as developed in Sections A.2 and A.3.

(ii) The "$O(1)$" constants that appear in the height machine (Theorem B.3.2) depend on the varieties, divisors, and morphisms, but they are independent of the points on the varieties. In principle, it is possible to construct all of the $h_{V,D}$'s explicitly and to give bounds for the $O(1)$'s in terms of the coefficients of the defining equations of the V's, D's, and ϕ's. Thus the height machine is effective. However, it is often difficult in practice to bound the $O(1)$'s. And even when bounds are calculated, the results are generally quite large, because explicit bounds usually depend on some explicit rendition of the Nullstellensatz, and this in turn requires the use of generalized resultants or elimination theory.

(iii) We have described the height machine for varieties defined over number fields. The same construction works more generally over any field with a proper set of absolute values satisfying the product formula. For example, there is a theory of heights over function fields. See Lang [6] for the general formulation.

PROOF. We construct the height machine in pieces. First, for every divisor $D \in \mathrm{Div}(V)$ whose linear system has no base points, we choose a morphism $\phi_D : V \to \mathbb{P}^n$ associated to D and define

$$h_{V,D}(P) = h\big(\phi_D(P)\big) \qquad \text{for all } P \in V(\bar{k}).$$

(That is, ϕ_D is a morphism such that $\phi_D^* H \sim D$ for any hyperplane H in $\mathbb{P}^n$.) Next, for every other divisor $D \in \mathrm{Div}(V)$, we use (A.3.2.3) to write D as a difference of divisors whose linear system has no base points, say $D = D_1 - D_2$ (we could even require D_1, D_2 to be very ample divisors), and then we define

$$h_{V,D}(P) = h_{V,D_1}(P) - h_{V,D_2}(P) \qquad \text{for all } P \in V(\bar{k}).$$

This gives us a height function $h_{V,D}$ for every divisor D on every variety V.

We begin by verifying that up to $O(1)$, the height function $h_{V,D}$ associated to a base point free divisor D is independent of the morphism ϕ_D. So let $\psi_D : V \to \mathbb{P}^m$ be another morphism associated to D. This means that $\phi_D^* H \sim D \sim \psi_D^* H'$, where H is a hyperplane in $\mathbb{P}^n$ and H' is a hyperplane in $\mathbb{P}^m$. Now Theorem B.3.1 tells us that

$$h\big(\phi_D(P)\big) = h\big(\psi_D(P)\big) + O(1) \qquad \text{for all } P \in V(\bar{k}).$$

Hence for base point free divisors, we can use any associated morphism to compute the height.

We next check additivity property (c) for base point free divisors, which we then use to show that the height $h_{V,D}$ is well-defined up to $O(1)$, independent of the decomposition $D = D_1 - D_2$.

Let then D and E be base point free divisors, and let $\phi_D : V \to \mathbb{P}^n$ and $\phi_E : V \to \mathbb{P}^m$ be associated morphisms. Composing the product $\phi_D \times \phi_E : V \to \mathbb{P}^n \times \mathbb{P}^m$ with the Segre embedding $S_{n,m}$ (A.1.2.6(b)) gives a morphism

$$\phi_D \otimes \phi_E : V \longrightarrow \mathbb{P}^N, \qquad \phi_D \otimes \phi_E(P) = S_{n,m}\big(\phi_D(P), \phi_E(P)\big).$$

The morphism $\phi_D \otimes \phi_E$ is associated to the divisor $D + E$, that is,

$$(\phi_D \otimes \phi_E)^* H \sim D + E,$$

see (B.2.4(a)). We showed above that the height for a base point free divisor can be computed using any associated morphism, so

$$h_{V,D+E}(P) = h\big((\phi_D \otimes \phi_E)(P)\big) + O(1) \qquad \text{for all } P \in V(\bar{k}).$$

Now we can use (B.2.4(b)) to obtain

$$\begin{aligned} h_{V,D+E}(P) &= h\big((\phi_D \otimes \phi_E)(P)\big) + O(1) \\ &= h\big(S_{n,m}\big(\phi_D(P), \phi_E(P)\big)\big) + O(1) \\ &= h\big(\phi_D(P)\big) + h\big(\phi_E(P)\big) + O(1) \quad \text{(from (B.2.4(b)))} \\ &= h_{V,D}(P) + h_{V,E}(P) + O(1). \end{aligned}$$

This gives additivity for base point free divisors. Suppose now that we have two decompositions

$$D = D_1 - D_2 = E_1 - E_2$$

of a divisor D as the difference of base point divisors. Then $D_1 + E_2 = D_2 + E_1$, and hence

$$\begin{aligned} h_{V,D_1} + h_{V,E_2} &= h_{V,D_1+E_2} + O(1) \\ &= h_{V,D_2+E_1} + O(1) \\ &= h_{V,D_2} + h_{V,E_1} + O(1), \end{aligned}$$

which implies that $h_{V,D_1} - h_{V,D_2} = h_{V,E_1} - h_{V,E_2} + O(1)$.

It is now easy to check properties (a) and (b). Thus if H is a hyperplane in $\mathbb{P}^n$, then the identity map $\mathbb{P}^n \to \mathbb{P}^n$, $P \mapsto P$, is associated to H. This gives (a). To verify (b), we write $D \in \operatorname{Div}(W)$ as a difference of base point free divisors, $D = D_1 - D_2$, and let ϕ_{D_1} and ϕ_{D_2} be the corresponding morphisms of W into projective space. Then $\phi^* D_1$ and $\phi^* D_2$ are base point free, with associated morphisms $\phi_{D_1} \circ \phi$ and $\phi_{D_2} \circ \phi$, respectively. Hence

$$\begin{aligned} h_{V,\phi^* D} &= h_{V,\phi^* D_1} - h_{V,\phi^* D_2} + O(1) \\ &= h \circ \phi_{D_1} \circ \phi - h \circ \phi_{D_2} \circ \phi + O(1) \\ &= h_{W,D_1} \circ \phi - h_{W,D_2} \circ \phi + O(1) \\ &= h_{W,D} \circ \phi + O(1). \end{aligned}$$

Next we check the additivity property (c), which we already know for base point free divisors. Now let D and E be arbitrary divisors, and write them as differences $D = D_1 - D_2$ and $E = E_1 - E_2$ of base point free (or even very ample) divisors. Then $D_1 + E_1$ and $D_2 + E_2$ are base point free, so we can compute

$$\begin{aligned} h_{V,D+E} &= h_{V,D_1+E_1} - h_{V,D_2+E_2} + O(1) \\ &= h_{V,D_1} + h_{V,E_1} - h_{V,D_2} - h_{V,E_2} + O(1) \\ &= h_{V,D} + h_{V,E} + O(1). \end{aligned}$$

This completes the proof of additivity (c).

We also note at this point that the normalization property (a), the functoriality property (b) for embeddings to projective space, and the additivity property (c) determine the height functions up to $O(1)$. The point is that if D is very ample with associated embedding $\phi_D : V \hookrightarrow \mathbb{P}^n$, then (a) and (b) imply that $h_{V,D} = h \circ \phi_D + O(1)$. This determines the height function for very ample divisors. But any divisor D can be written as the difference $D_1 - D_2$ of very ample divisors (A.3.2.3), so the additivity (c) forces us to define $h_{V,D} = h_{V,D_1} - h_{V,D_2} + O(1)$. This proves the uniqueness property (h) of the height.

Next suppose that D and E are linearly equivalent. Writing $D = D_1 - D_2$ and $E = E_1 - E_2$ as the difference of base point free divisors as usual, we have $D_1 + E_2 \sim D_2 + E_1$. This means that the morphisms $\phi_{D_1+E_2}$ and $\phi_{D_2+E_1}$ are associated to the same linear system, so Theorem B.3.1 (or even Theorem B.2.5) tells us that

$$h\big(\phi_{D_1+E_2}(P)\big) = h\big(\phi_{D_2+E_1}(P)\big) + O(1) \qquad \text{for all } P \in V(\bar{k}).$$

Using this equality and additivity gives

$$h_{V,D_1} + h_{V,E_2} = h_{V,D_1+E_2} + O(1) = h_{V,D_2+E_1} + O(1) = h_{V,D_2} + h_{V,E_1} + O(1).$$

Hence

$$h_{V,D} = h_{V,D_1} - h_{V,D_2} + O(1) = h_{V,E_1} - h_{V,E_2} + O(1) = h_{V,E} + O(1),$$

which proves (d).

To prove positivity (e), we take $D > 0$ and write $D = D_1 - D_2$ as a difference of base point free divisors as usual. Choose a basis $f_0, \ldots, f_n$ for $L(D_2)$. Then the fact that D is positive implies that

$$D_1 + \operatorname{div}(f_i) = D + D_2 + \operatorname{div}(f_i) \geq 0,$$

so $f_0, \ldots, f_n$ are also in $L(D_1)$. We extend this set to form a basis

$$f_0, \ldots, f_n, f_{n+1}, \ldots, f_m \in L(D_1).$$

These bases give us morphisms

$$\phi_{D_1} = (f_0, \ldots, f_m) : V \longrightarrow \mathbb{P}^m \qquad \text{and} \qquad \phi_{D_2} = (f_0, \ldots, f_n) : V \longrightarrow \mathbb{P}^n$$

associated to D_1 and D_2. The functions $f_0, \ldots, f_m$ are regular at all points not in the support of D_1, so for any $P \in V$ with $P \notin \operatorname{supp}(D_1)$ we can compute

$$\begin{aligned} h_{V,D}(P) &= h_{V,D_1}(P) - h_{V,D_2}(P) + O(1) \\ &= h\big(\phi_{D_1}(P)\big) - h\big(\phi_{D_2}(P)\big) + O(1) \\ &= h\big(f_0(P), \ldots, f_m(P)\big) - h\big(f_0(P), \ldots, f_n(P)\big) + O(1) \\ &\geq O(1). \end{aligned}$$

The last line follows directly from the definition of the height, since the fact that $m \geq n$ clearly implies

$$\prod_{v \in M_k} \max_{0 \leq i \leq m} \big\{ \|f_i(P)\|_v \big\} \geq \prod_{v \in M_k} \max_{0 \leq i \leq n} \big\{ \|f_i(P)\|_v \big\}.$$

This gives the desired estimate for points not in the support of D_1. Now choose very ample divisors $H_0, H_1, \ldots, H_r$ on V with the property that $H_0 \cap \cdots \cap H_r = \emptyset$ and $H_i + D$ is very ample. For example, use (A.3.2.3) to find a very ample divisor H such that $D + H$ is also very ample, take an embedding $V \hookrightarrow \mathbb{P}^r$ corresponding to H, and take the H_i's to be the pullbacks of the coordinate hyperplanes in $\mathbb{P}^r$. Now we apply our above result to each of the decompositions $D = (D + H_i) - H_i$ to deduce the inequality $h_{V,D} \geq O(1)$ for all points not in the support of D. Finally, varying D in its linear system $|D|$, we obtain the positivity property (e) for all points not lying in the base locus of $|D|$.

We will give a proof of the algebraic equivalence (f) using the fact that if D is ample and E is algebraically equivalent to 0, then there is an integer $m > 0$ such that $mD + nE$ is base point free for all integers n. (See Lang [6, Chapter 4, Lemma 3.2].) However, we will later (Theorem B.5.9) prove a stronger result, using the theory of canonical heights and functorial properties of the Picard and Albanese varieties.

The height associated to a base point free divisor is nonnegative by construction (or by the positivity property (e) with empty base locus), so

$$h_{V,mD+nE}(P) \geq O(1) \qquad \text{for all } P \in V(\bar{k}).$$

Using additivity (c), we obtain

$$mh_{V,D}(P) + nh_{V,E}(P) \geq -c \qquad \text{for all } P \in V(\bar{k}),$$

where the constant $-c$ will depend on D, E, m, and n, but is independent of P. This holds for all integers n, so we can rewrite using positive and negative values for n. Thus for any $n \geq 1$ we obtain

$$\frac{m}{n} + \frac{c}{nh_{V,D}(P)} \geq \frac{h_{V,E}(P)}{h_{V,D}(P)} \geq -\frac{m}{n} - \frac{c}{nh_{V,D}(P)} \qquad \text{for all } P \in V(\bar{k}).$$

It is important to keep in mind that the constant c depends on n. We now let $h_{V,D}(P) \to \infty$. This destroys the c's and yields

$$\frac{m}{n} \geq \limsup_{h_{V,D}(P)\to\infty} \frac{h_{V,E}(P)}{h_{V,D}(P)} \geq \liminf_{h_{V,D}(P)\to\infty} \frac{h_{V,E}(P)}{h_{V,D}(P)} \geq -\frac{m}{n}.$$

These inequalities hold for all $n \geq 1$, so letting $n \to \infty$, we obtain the desired result,

$$\lim_{h_{V,D}(P)\to\infty} \frac{h_{V,E}(P)}{h_{V,D}(P)} = 0.$$

This completes the proof of the algebraic equivalence property (f).

It remains to prove the finiteness property (g). Note that if we replace the ample divisor D by a very ample multiple mD, then additivity (c)

implies that $h_{V,mD} = mh_{V,D} + O(1)$; hence it suffices to prove (g) under the assumption that D is very ample. Let $\phi : V \hookrightarrow \mathbb{P}^n$ be an embedding associated to D, so $\phi^*H = D$. Then (a) and (b) imply that

$$h_{V,D} \circ \phi = h_{\mathbb{P}^n,\phi^*D} + O(1) = h_{\mathbb{P}^n,H} + O(1) = h + O(1),$$

so we are reduced to showing that $\mathbb{P}^n(k')$ has finitely many points of bounded height. This follows from (B.2.3), which completes the proof of (g), and with it the proof of Theorem B.3.2. □

Remark B.3.3. We illustrate the use of the height machine by quickly reproving a special case of Theorem B.2.5. Let $\phi : \mathbb{P}^n \to \mathbb{P}^m$ be a morphism of degree d, and let H_n and H_m be hyperplanes in $\mathbb{P}^n$ and $\mathbb{P}^m$, respectively. The assumption that ϕ has degree d means that $\phi^*H_m \sim dH_n$. For any $P \in \mathbb{P}^n(\bar{\mathbb{Q}})$ we compute

$$\begin{aligned}
h(\phi(P)) &= h_{\mathbb{P}^m,H_m}(\phi(P)) + O(1) && \text{from B.3.2(a)}\\
&= h_{\mathbb{P}^n,\phi^*H_m}(P) + O(1) && \text{from B.3.2(b)}\\
&= h_{\mathbb{P}^n,dH_n}(P) + O(1) && \text{from B.3.2(d) (note } \phi^*H_m \sim dH_n)\\
&= dh_{\mathbb{P}^n,H_n}(P) + O(1) && \text{from B.3.2(c)}\\
&= dh(P) + O(1) && \text{from B.3.2(a) again.}
\end{aligned}$$

As another illustration of the power of the height machine, we use the divisor class relations from Section A.7.2 to derive some important height formulas on abelian varieties. These formulas and the finiteness result (B.3.2(g)) will provide half of the necessary tools for applying the descent theorem (C.0.3) to abelian varieties. More precisely, the weak Mordell–Weil theorem (C.0.2) says that the group $A(k)/mA(k)$ is finite. Then (B.3.2(g)) and (B.3.4(a,b)) will be used to deduce that the group $A(k)$ itself is finitely generated.

Corollary B.3.4. *Let A/k be an abelian variety defined over a number field, and let $D \in \operatorname{Div}(A)$ be a divisor on A.*
(a) *Let m be an integer. Then for all $P \in A(\bar{k})$,*

$$h_{A,D}([m]P) = \frac{m^2+m}{2} h_{A,D}(P) + \frac{m^2-m}{2} h_{A,D}(-P) + O(1).$$

In particular, if D has a symmetric divisor class (i.e., $[-1]^*D \sim D$), then

$$h_{A,D}([m]P) = m^2 h_{A,D}(P) + O(1),$$

and if D has an antisymmetric divisor class (i.e., $[-1]^*D \sim -D$), then

$$h_{A,D}([m]P) = mh_{A,D}(P) + O(1).$$

(Note that the $O(1)$'s depend on A, D, and m.)

(b) *If D has a symmetric divisor class, then for all $P, Q \in A(\bar{k})$,*

$$h_{A,D}(P+Q) + h_{A,D}(P-Q) = 2h_{A,D}(P) + 2h_{A,D}(Q) + O(1).$$

(c) *If D has an antisymmetric divisor class, then*

$$h_{A,D}(P+Q) = h_{A,D}(P) + h_{A,D}(Q) + O(1) \qquad \text{for all } P, Q \in A(\bar{k}).$$

PROOF. (a) Mumford's formula (A.7.2.5) tells us that

$$[m]^*D \sim \frac{m^2+m}{2}D + \frac{m^2-m}{2}[-1]^*D.$$

Using this and standard properties of heights (B.3.2(b),(d),(c)), we obtain a corresponding height relation

$$\begin{aligned} h_{A,D}([m]P) &= h_{A,[m]^*D}(P) + O(1) \\ &= h_{A,(1/2)(m^2+m)D+(1/2)(m^2-m)[-1]^*D}(P) + O(1) \\ &= \frac{m^2+m}{2}h_{A,D}(P) + \frac{m^2-m}{2}h_{A,[-1]^*D}(P) + O(1) \\ &= \frac{m^2+m}{2}h_{A,D}(P) + \frac{m^2-m}{2}h_{A,D}(-P) + O(1). \end{aligned}$$

This proves the first part of (a). The other two parts are consequences of the relation $h_{A,D} \circ [-1] = \pm h_{A,D} + O(1)$, where the sign is positive (respectively negative) if the divisor class of D is symmetric (respectively antisymmetric).

(b) Consider the following four maps from $A \times A$ to A:

$$\sigma, \delta, \pi_1, \pi_2 : A \times A \longrightarrow A, \qquad \begin{cases} \sigma(P,Q) = P+Q \\ \delta(P,Q) = P-Q \\ \pi_1(P,Q) = P \\ \pi_2(P,Q) = Q. \end{cases}$$

Proposition A.7.3.3 gives the divisor class relation

$$\sigma^*D + \delta^*D \sim 2\pi_1^*D + 2\pi_2^*D$$

on $A \times A$. We use this divisor relation and standard properties of height functions (B.3.2) to compute

$$\begin{aligned} h_{A,\sigma^*D}(P,Q) + h_{A,\delta^*D}(P,Q) &= 2h_{A,\pi_1^*D}(P,Q) + 2h_{A,\pi_2^*D}(P,Q) + O(1), \\ h_{A,D}(\sigma(P,Q)) + h_{A,D}(\delta(P,Q)) &= 2h_{A,D}(\pi_1(P,Q)) \\ &\qquad + 2h_{A,D}(\pi_2(P,Q)) + O(1), \\ h_{A,D}(P+Q) + h_{A,D}(P-Q) &= 2h_{A,D}(P) + 2h_{A,D}(Q) + O(1). \end{aligned}$$

(c) Let σ, π_1, π_2 be as above. Then Proposition A.7.3.2 gives the formula

$$\sigma^* D \sim \pi_1^* D + \pi_2^* D.$$

Again we use (B.3.2) to translate this divisor relation into a height relation,

$$\begin{aligned} h_{A,\sigma^* D}(P,Q) &= h_{A,\pi_1^* D}(P,Q) + h_{A,\pi_2^* D}(P,Q) + O(1), \\ h_{A,D}\bigl(\sigma(P,Q)\bigr) &= h_{A,D}\bigl(\pi_1(P,Q)\bigr) + h_{A,D}\bigl(\pi_2(P,Q)\bigr) + O(1), \\ h_{A,D}(P+Q) &= h_{A,D}(P) + h_{A,D}(Q) + O(1). \end{aligned}$$ □

The following result will be used in Part D to study integer points on curves. The first part of (B.3.5) is an immediate consequence of the algebraic equivalence property of the height machine (B.3.2(f)) and can be strengthened by using Theorem B.5.9 below, but we will give a direct proof, since algebraic equivalence on curves is much more elementary than in the general case.

Proposition B.3.5. *Let C/k be a smooth projective curve.*
(a) *Let $D, E \in \operatorname{Div}(C)$ be divisors with* $\deg(D) \geq 1$. *Then*

$$\lim_{\substack{P \in C(\bar{k}) \\ h_D(P) \to \infty}} \frac{h_E(P)}{h_D(P)} = \frac{\deg E}{\deg D}.$$

(b) *Let $f, g \in k(C)$ be rational functions on C with f nonconstant. Then*

$$\lim_{\substack{P \in C(\bar{k}) \\ h(f(P)) \to \infty}} \frac{h\bigl(g(P)\bigr)}{h\bigl(f(P)\bigr)} = \frac{\deg g}{\deg f}.$$

PROOF. Let $d = \deg(D)$ and $e = \deg(E)$. For every integer n, both positive and negative, consider the divisor

$$A_n = n(eD - dE) + D.$$

Notice that $\deg(A_n) = \deg(D) \geq 1$, so A_n is ample (A.4.2.4). The positivity property of the height machine (B.3.2e) implies that $h_{A_n}(P)$ is bounded below for all $P \in C(\bar{k})$. Or we can prove it directly as follows. If $A \in \operatorname{Div}(C)$ is any ample divisor, then mA is very ample for some $m \geq 1$, say mA is associated to the map $\phi : C \to \mathbb{P}^N$. Then

$$h_A(P) = \frac{1}{m} h_{mA}(P) + O(1) = \frac{1}{m} h\bigl(\phi(P)\bigr) + O(1) \geq O(1),$$

since the height on $\mathbb{P}^N(\bar{k})$ is nonnegative.

We now know that h_{A_n} is bounded below on $C(\bar{k})$, so using the additivity of the height, we find that

$$O(1) \le h_{A_n}(P) = h_{n(eD-dE)+D}(P) = n\bigl(eh_D(P) - dh_E(P)\bigr) + h_D(P) \qquad \text{for all } P \in C(\bar{k}).$$

Of course, we must not forget that the $O(1)$ may depend on D, E, and n, so we will denote it by $-\kappa(D,E,n)$. Rearranging our inequality, we find that

$$\frac{-\kappa(D,E,n)}{dh_D(P)} \le n\left(\frac{e}{d} - \frac{h_E(P)}{h_D(P)}\right) + \frac{1}{d} \qquad \text{for all } P \in C(\bar{k}).$$

This holds for positive and negative values of n, so taking both n and $-n$ with $n \ge 1$, we obtain the estimate

$$\frac{-\kappa(D,E,n)}{ndh_D(P)} - \frac{1}{nd} \le \frac{e}{d} - \frac{h_E(P)}{h_D(P)} \le \frac{\kappa(D,E,-n)}{ndh_D(P)} + \frac{1}{nd} \qquad \text{for all } P \in C(\bar{k}).$$

Now consider what happens as $h_D(P) \to \infty$. We find that

$$-\frac{1}{nd} \le \liminf_{h_D(P)\to\infty} \left(\frac{e}{d} - \frac{h_E(P)}{h_D(P)}\right) \le \limsup_{h_D(P)\to\infty} \left(\frac{e}{d} - \frac{h_E(P)}{h_D(P)}\right) \le \frac{1}{nd}.$$

These inequalities hold for all $n \ge 1$, so we can let $n \to \infty$ to obtain

$$\lim_{h_D(P)\to\infty} \frac{e}{d} - \frac{h_E(P)}{h_D(P)} = 0.$$

This completes the proof of (a).
(b) This follows easily from (a). Write $\operatorname{div}(f) = D' - D$ and $\operatorname{div}(g) = E' - E$. Note that $\deg(f) = \deg(D)$ and $\deg(g) = \deg(E)$. Further, if we consider f to be a map $f : C \to \mathbb{P}^1$, then $D = f^*(\infty)$, so $h_D = h \circ f + O(1)$. Similarly, $h_E = h \circ g + O(1)$. Now we use (a) to compute

$$\lim_{h(f(P))\to\infty} \frac{h\bigl(g(P)\bigr)}{h\bigl(f(P)\bigr)} = \lim_{h_D(P)\to\infty} \frac{h_E(P) + O(1)}{h_D(P) + O(1)} = \frac{\deg E}{\deg D} = \frac{\deg g}{\deg f}.$$

□

Up to a bounded function, the height $h_{V,D}$ associated to a divisor D depends only on the divisor class of D. It is sometimes convenient to reformulate the height machine (B.3.2) purely in terms of divisor classes or line bundles.

Theorem B.3.6. *Let V be a projective variety defined over a number field k. There is a unique homomorphism*

$$h_V : \operatorname{Pic}(V) \longrightarrow \frac{\{\text{functions } V(\bar{k}) \to \mathbb{R}\}}{\{\text{bounded functions } V(\bar{k}) \to \mathbb{R}\}}$$

with the property that if $\mathcal{L} \in \operatorname{Pic}(V)$ is very ample and $\phi_{\mathcal{L}} : V \hookrightarrow \mathbb{P}^n$ is an associated embedding, then

$$h_{V,\mathcal{L}} = h \circ \phi_{\mathcal{L}} + O(1).$$

The height functions $h_{V,\mathcal{L}}$ have the following additional properties:
(a) (Functoriality) *Let $\phi : V \to W$ be a morphism of smooth varieties, and let $\mathcal{L} \in \operatorname{Pic}(W)$. Then*

$$h_{V,\phi^*\mathcal{L}} = h_{W,\mathcal{L}} \circ \phi + O(1).$$

(b) (Positivity) *Let B be the base locus of $\mathcal{L} \in \operatorname{Pic}(V)$, and assume that $B \neq V$. Then*

$$h_{V,\mathcal{L}} \geq O(1) \qquad \text{on } V \smallsetminus B.$$

(c) (Algebraic Equivalence) *Let $\mathcal{L}, \mathcal{M} \in \operatorname{Pic}(V)$ with $\mathcal{L}$ ample and $\mathcal{M}$ algebraically equivalent to zero. Then*

$$\lim_{\substack{P \in V(\bar{k}) \\ h_{V,\mathcal{L}}(P) \to \infty}} \frac{h_{V,\mathcal{M}}(P)}{h_{V,\mathcal{L}}(P)} = 0.$$

(See Theorem B.5.9 for a stronger statement.)

PROOF. All of this is a restatement, in terms of line bundles, of the height machine (B.3.2). Note that the linear equivalence and additivity properties of (B.3.2) are included in the statement that the height mapping h_V is defined and is a homomorphism on $\operatorname{Pic}(V)$ and that we do not need a smoothness hypotheses because $\operatorname{Pic}(V)$ is defined in terms of Cartier divisors.

□

B.4. Canonical Height Functions

The height machine (B.3.2) associates to each divisor $D \in \mathrm{Div}(V)$ a height function $h_D : V(\bar{k}) \to \mathbb{R}$. These height functions are well-defined and satisfy various properties modulo $O(1)$. In some cases it is possible to find a particular height function within its $O(1)$ equivalence class that has particularly nice properties. This theory, which was developed by Néron and Tate, will form the subject of this and the next section.

Theorem B.4.1. (Néron, Tate) *Let V/k be a smooth variety defined over a number field, let $D \in \mathrm{Div}(V)$, and let $\phi : V \to V$ be a morphism. Suppose that*

$$\phi^* D \sim \alpha D$$

for some number $\alpha > 1$. Then there is a unique function, called the canonical height on V relative to ϕ and D,

$$\hat{h}_{V,\phi,D} : V(\bar{k}) \longrightarrow \mathbb{R},$$

with the following two properties:
(i) $\hat{h}_{V,\phi,D}(P) = h_{V,D}(P) + O(1)$ *for all* $P \in V(\bar{k})$.
(ii) $\hat{h}_{V,\phi,D}\big(\phi(P)\big) = \alpha \hat{h}_{V,\phi,D}(P)$ *for all* $P \in V(\bar{k})$.

The canonical height depends only on the linear equivalence class of D. Further, it can be computed as the limit

$$\hat{h}_{V,\phi,D}(P) = \lim_{n\to\infty} \frac{1}{\alpha^n} h_{V,D}\big(\phi^n(P)\big),$$

where $\phi^n = \phi \circ \phi \circ \cdots \circ \phi$ is the n-fold iterate of ϕ.

Proof. Applying the height machine to the relation $\phi^* D \sim \alpha D$, we find that there is a constant C such that

$$\big|h_{V,D}\big(\phi(Q)\big) - \alpha h_{V,D}(Q)\big| \le C \qquad \text{for all } Q \in V(\bar{k}).$$

(N.B. C depends on V, D, ϕ, and the choice of the height function $h_{V,D}$.) Now take any point $P \in V(\bar{k})$. We are going to prove that the sequence $\alpha^{-n} h_{V,D}\big(\phi^n(P)\big)$ converges by verifying that it is Cauchy. We take $n \ge m$ and compute

$$\begin{aligned}
&\big|\alpha^{-n} h_{V,D}\big(\phi^n(P)\big) - \alpha^{-m} h_{V,D}\big(\phi^m(P)\big)\big| \\
&\qquad = \left| \sum_{i=m+1}^{n} \alpha^{-i}\Big(h_{V,D}\big(\phi^i(P)\big) - \alpha h_{V,D}\big(\phi^{i-1}(P)\big)\Big)\right| \\
&\qquad\qquad\qquad\qquad \text{(telescoping sum)}
\end{aligned}$$

$$\leq \sum_{i=m+1}^{n} \alpha^{-i} \left| h_{V,D}\big(\phi^i(P)\big) - \alpha h_{V,D}\big(\phi^{i-1}(P)\big) \right|$$

(triangle inequality)

$$\leq \sum_{i=m+1}^{n} \alpha^{-i} C \quad \text{(from above with } Q = \phi^{i-1}P\text{)}$$

$$= \left(\frac{\alpha^{-m} - \alpha^{-n}}{\alpha - 1} \right) C.$$

This last quantity goes to 0 as $n > m \to \infty$, which proves that the sequence is Cauchy, hence converges. So we can define $\hat{h}_{V,\phi,D}(P)$ to be the limit

$$\hat{h}_{V,\phi,D}(P) = \lim_{n\to\infty} \frac{1}{\alpha^n} h_{V,D}\big(\phi^n(P)\big).$$

To verify property (i), we take $m = 0$ and let $n \to \infty$ in the inequality proven above. This gives

$$\left| \hat{h}_{V,\phi,D}(P) - h_{V,D}(P) \right| \leq \frac{C}{\alpha - 1},$$

which is an explicit form of the desired estimate. Property (ii) follows directly from the definition of the canonical height as a limit, since once we know that the limit exists we can compute

$$\begin{aligned} \hat{h}_{V,\phi,D}\big(\phi(P)\big) &= \lim_{n\to\infty} \frac{1}{\alpha^n} h_{V,D}\big(\phi^n(\phi(P))\big) \\ &= \lim_{n\to\infty} \frac{\alpha}{\alpha^{n+1}} h_{V,D}\big(\phi^{n+1}(P)\big) \\ &= \alpha \hat{h}_{V,\phi,D}(P). \end{aligned}$$

Finally, in order to prove the uniqueness, suppose that $\hat{h}$ and $\hat{h}'$ are two functions with properties (i) and (ii). Let $g = \hat{h} - \hat{h}'$ be the difference. Then (i) implies that g is bounded, say $|g(P)| \leq C'$ for all $P \in V(\bar{k})$. On the other hand, (ii) says that $g \circ \phi = \alpha g$, and iterating this relation gives $g \circ \phi^n = \alpha^n g$ for all $n \geq 1$. Hence

$$|g(P)| = \frac{\big|g\big(\phi^n(P)\big)\big|}{\alpha^n} \leq \frac{C'}{\alpha^n} \xrightarrow[n\to\infty]{} 0.$$

This proves that $g(P) = 0$ for all P, so $\hat{h} = \hat{h}'$. □

To understand the condition that $\phi^* D$ be linearly equivalent to αD, we observe that ϕ induces a $\mathbb{Z}$-linear map $\phi^* : \operatorname{Pic}(V) \to \operatorname{Pic}(V)$. Tensoring with $\mathbb{R}$ gives a linear transformation

$$\phi^* : \operatorname{Pic}(V) \otimes \mathbb{R} \longrightarrow \operatorname{Pic}(V) \otimes \mathbb{R}$$

of $\mathbb{R}$-vector spaces. In order to apply the averaging process of Theorem B.4.1 to construct a canonical height, we need a divisor class that is an eigenvector for this linear transformation and that has an eigenvalue strictly greater than 1. For example, any morphism $\phi : \mathbb{P}^n \to \mathbb{P}^n$ of degree n has the property that $\phi^*H \sim nH$. Similarly, if D is a symmetric divisor on an abelian variety A, then the multiplication map $[n] : A \to A$ satisfies $[n]^*D \sim n^2D$. We will discuss these examples further below. For an example of canonical heights defined on certain K3 surfaces, see Silverman [6].

If the divisor D is ample, then the canonical height can be used to prove various finiteness results. It is convenient to use some terminology from the theory of dynamical systems.

Definition. Let S be a set, let $\phi : S \to S$ be a function, and for each $n \geq 1$ let $\phi^n : S \to S$ denote the nth iterate of ϕ. An element $P \in S$ is called *periodic for* ϕ if $\phi^n(P) = P$ for some $n \geq 1$, and it is called *preperiodic for* ϕ if $\phi^n(P)$ is periodic for some $n \geq 1$. Equivalently, P is preperiodic if its *forward orbit*

$$O_\phi^+(P) = \{P, \phi(P), \phi^2(P), \phi^3(P), \ldots\}$$

is finite.

Proposition B.4.2. *Let $\phi : V \to V$ be a morphism of a variety defined over a number field k. Let $D \in \operatorname{Div}(V)$ be an ample divisor such that $\phi^*D \sim \alpha D$ for some $\alpha > 1$, and let $\hat{h}_{V,\phi,D}$ be the associated canonical height (B.4.1).*
(a) *Let $P \in V(\bar{k})$. Then $\hat{h}_{V,\phi,D}(P) \geq 0$, and*

$$\hat{h}_{V,\phi,D}(P) = 0 \iff P \text{ is preperiodic for } \phi.$$

(b) (Northcott [3]) *The set*

$$\{P \in V(k) \mid P \text{ is preperiodic for } \phi\}$$

is finite.

PROOF. (a) The fact that D is ample means that we can choose a height function $h_{V,D}$ with nonnegative values. It is then immediate from the definition that $\hat{h}_{V,\phi,D}$ is nonnegative.

Now let $P \in V(\bar{k})$. Replacing k by a finite extension, we may assume that $P \in V(k)$ and that D and ϕ are defined over k. Suppose first that P is preperiodic for ϕ. Then the sequence $(\phi^n P)_{n\geq 1}$ repeats, so the sequence of heights $\big(h_{V,D}(\phi^n P)\big)_{n\geq 1}$ is bounded. It follows that $\alpha^{-n}h_{V,D}(\phi^n P) \to 0$ as $n \to \infty$, so Theorem B.4.1 says that $\hat{h}_{V,\phi,D}(P) = 0$.

Conversely, suppose that $\hat{h}_{V,\phi,D}(P) = 0$. Then for any $n \geq 1$ we have

$$h_{V,D}(\phi^n P) = \hat{h}_{V,\phi,D}(\phi^n P) + O(1) = \alpha^n \hat{h}_{V,\phi,D}(P) + O(1) = O(1).$$

Note further that all of the points $\phi^n P$ are in $V(k)$. Thus there is a constant B such that

$$O_\phi^+(P) = \{P, \phi(P), \phi^2(P), \phi^3(P), \ldots\} \subset \{Q \in V(k) \mid h_{V,D}(Q) \leq B\}.$$

But D is ample, so Theorem B.3.2(g) says that there are only finitely many points in $V(k)$ with bounded $h_{V,D}$-height. Hence $O_\phi^+(P)$ is finite, so P is preperiodic for ϕ. □

Remarks B.4.3. Two important cases to which we can apply Proposition B.4.2 are projective spaces and abelian varieties.

(a) Let $\phi : \mathbb{P}^n \to \mathbb{P}^n$ be a morphism of degree $d \geq 2$. Then $\phi^* H \sim dH$ for any hyperplane $H \in \operatorname{Div}(\mathbb{P}^n)$. It follows from (B.4.2b) that the set

$$\{P \in \mathbb{P}^n(k) \mid P \text{ is preperiodic for } \phi\}$$

is finite. This result, which is due to Northcott [3], can also be proven directly using Theorem B.2.5.

As a special case, consider the map $\phi(x_0, \ldots, x_n) = (x_0^2, \ldots, x_n^2)$. It is easy to see that the canonical height associated to ϕ is simply the usual "noncanonical" height function on $\mathbb{P}^n$. Further, a point is preperiodic for ϕ if and only if all of the (defined) ratios x_j/x_i are roots of unity or zero. Thus this special case of (B.4.2a) is Kronecker's theorem (B.2.3.1).

(b) Let A be an abelian variety, let $D \in \operatorname{Div}(A)$ be an ample symmetric divisor, and let $[n] : A \to A$ be the multiplication-by-n map for some $n \geq 2$. Then $[n]^* D \sim n^2 D$ from (A.7.2.5), so we can apply (B.4.2) to conclude that $A(k)$ has only finitely many points that are preperiodic for $[n]$. But $P \in A$ is preperiodic for $[n]$ if and only if there are integers $i > j$ such that $[n^i]P = [n^j]P$, so we see that the $[n]$-preperiodic points are precisely the torsion points. Hence $A(k)_{\text{tors}}$ is finite. This global proof of the finiteness of $A(k)_{\text{tors}}$ may be compared with the local proof (see Theorem C.1.4 and the remark following it). The local proof uses the fact that for all but finitely many primes $\mathfrak{p}$, the prime-to-p torsion in $A(k)$ injects when we reduce modulo $\mathfrak{p}$, where p is the residue characteristic of $\mathfrak{p}$.

B.5. Canonical Heights on Abelian Varieties

The construction of Néron and Tate (B.4.1) associates a canonical height to any morphism $\phi : V \to V$ with an eigendivisor $\phi^*D \sim \alpha D$ having eigenvalue $\alpha > 1$. An important example is the case of an abelian variety A, a symmetric divisor D, and a multiplication-by-m map $[m] : A \to A$, since Corollary A.7.2.5 tells us that $[m]^*D \sim m^2 D$. It turns out that the resulting canonical height is independent of the choice of $m \geq 2$ and has many additional useful properties. In particular, it is a quadratic form relative to the group law on A.

Similarly, if we choose an antisymmetric divisor D on A (i.e., D satisfies $[-1]^*D \sim D$), then (A.7.2.5) says that $[m]^*D \sim mD$, so again we get a canonical height. Finally, for any D we can write $2D$ as the sum of a symmetric divisor and an antisymmetric divisor, so using linearity we get a canonical height for any D.

We begin this section by describing the canonical heights associated to symmetric divisors. These are the heights that are most often used in Diophantine applications. Then at the end of the section we will develop the theory of canonical heights for arbitrary divisors and use it to describe the height pairing on an abelian variety and its dual. We conclude the section by applying the theory of canonical heights to deduce a strong form of the algebraic equivalence property (Theorem B.3.2(f)) for (not necessarily abelian) varieties.

Theorem B.5.1. (Néron, Tate) *Let A/k be an abelian variety defined over a number field, and let $D \in \mathrm{Div}(A)$ be a divisor whose divisor class is symmetric (i.e., $[-1]^*D \sim D$). There is a height function*

$$\hat{h}_{A,D} : A(\bar{k}) \longrightarrow \mathbb{R},$$

called the canonical height on A relative to D, with the following properties:

(a)

$$\hat{h}_{A,D}(P) = h_{A,D}(P) + O(1) \qquad \text{for all } P \in A(\bar{k}).$$

(b) *For all integers m,*

$$\hat{h}_{A,D}([m]P) = m^2 \hat{h}_{A,D}(P) \qquad \text{for all } P \in A(\bar{k}).$$

(c) *(Parallelogram Law)*

$$\hat{h}_{A,D}(P+Q)+\hat{h}_{A,D}(P-Q) = 2\hat{h}_{A,D}(P)+2\hat{h}_{A,D}(Q) \qquad \text{for all } P, Q \in A(\bar{k}).$$

(d) *The canonical height map* $\hat{h}_{A,D} : A(\bar{k}) \to \mathbb{R}$ *is a quadratic form. The associated pairing* $\langle \cdot , \cdot \rangle_D : A(\bar{k}) \times A(\bar{k}) \longrightarrow \mathbb{R}$ *defined by*

$$\langle P, Q \rangle_D = \frac{\hat{h}_{A,D}(P+Q) - \hat{h}_{A,D}(P) - \hat{h}_{A,D}(Q)}{2}$$

is bilinear and satisfies $\langle P, P \rangle = \hat{h}_{A,D}(P)$.
(e) (Uniqueness) *The canonical height* $\hat{h}_{A,D}$ *depends only on the divisor class of the divisor D. It is uniquely determined by (a) and (b) for any one integer* $m \geq 2$.

PROOF. We take $\hat{h}_{A,D}$ to be the canonical height on A with respect to the doubling map $[2] : A \to A$. Note that $[2]^*D \sim 4D$ from (A.7.2.5), so we can apply Theorem B.4.1 to obtain

$$\hat{h}_{A,D}(P) = \lim_{n\to\infty} \frac{1}{4^n} h_{A,D}\big([2^n]P\big).$$

Theorem B.4.1 tells us that $\hat{h}_{A,D} = h_{A,D} + O(1)$ and $\hat{h}_{A,D} \circ [2] = 4\hat{h}_{A,D}$, which gives (a) and also (b) for $m = 2$.

We could prove (b) by first proving (c) and then using induction, but we will give a direct proof. Corollary B.3.4(a) tells us that $h_{A,D}\big([m]Q\big) = m^2 h_{A,D}(Q) + O(1)$. This holds for all $Q \in A(\bar{k})$, where the $O(1)$ is bounded independently of Q. We replace Q by $[2]^n P$, divide by 4^n, and let $n \to \infty$. The result is

$$\begin{aligned} \hat{h}_{A,D}\big([m]P\big) &= \lim_{n\to\infty} \frac{1}{4^n} h_{A,D}\big([2^n m]P\big) \\ &= \lim_{n\to\infty} \frac{1}{4^n} \Big(m^2 h_{A,D}\big([2^n]P\big) + O(1)\Big) = m^2 \hat{h}_{A,D}(P). \end{aligned}$$

Notice how the limiting process eliminates the $O(1)$. Also note the crucial use made of the fact that the maps $[m]$ and $[2^n]$ commute with one another. This completes the proof of (b).

For (c), use the relation

$$h_{A,D}(P+Q) + h_{A,D}(P-Q) = 2h_{A,D}(P) + 2h_{A,D}(Q) + O(1)$$

from Corollary B.3.4(b). Note that the $O(1)$ is bounded independently of P and Q. Thus we can replace P and Q by $2^n P$ and $2^n Q$, divide by 4^n, and let $n \to \infty$. The $O(1)$ disappears, and we are left with the parallelogram law (c).

It is a standard fact that a function on an abelian group that satisfies the parallelogram law is a quadratic form. We will recall the proof below (B.5.2). This gives (d). Finally, the uniqueness statement (e) follows from the uniqueness assertion in Theorem B.4.1, since $\hat{h}_{A,D}$ is a canonical height relative to every map $[m]$. □

The following elementary result was used in proving Theorem B.5.1.

Lemma B.5.2. *Let A be an abelian group, and let $h : A \to \mathbb{R}$ be a function satisfying the parallelogram law,*

$$h(P+Q)+h(P-Q)=2h(P)+2h(Q) \qquad \text{for all } P,Q \in A.$$

Then h is a quadratic form on A.

PROOF. Putting $P = Q = 0$ into the parallelogram law gives $h(0) = 0$, and then putting $P = 0$ gives $h(-Q) = h(Q)$, so h is an even function. It remains to check that h is quadratic. We apply the parallelogram law four times (and use the evenness of h) to obtain

$$\begin{aligned} h(P+Q+R)+h(P+R-Q)-2h(P+R)-2h(Q)&=0,\\ h(P-R+Q)+h(P+R-Q)-2h(P)-2h(R-Q)&=0,\\ h(P-R+Q)+h(P+R+Q)-2h(P+Q)-2h(R)&=0,\\ 2h(R+Q)+2h(R-Q)-4h(R)-4h(Q)&=0. \end{aligned}$$

The alternating sum of these four equations is the desired result. □

If D is an ample symmetric divisor on A, then Proposition B.4.2 tells us that $\hat{h}_{A,D}(P) = 0$ if and only if P is a torsion point. Thus $\hat{h}_{A,D}$ is a positive definite quadratic form on $A(\bar{k})/(\text{torsion})$. The next result says that more is true.

Proposition B.5.3. *Let A/k be an abelian variety defined over a number field, and let $D \in \operatorname{Div}(A)$ be an ample divisor with symmetric divisor class.*
(a) *For all $P \in A(\bar{k})$, we have $\hat{h}_{A,D}(P) \ge 0$, with equality if and only if P is a point of finite order.*
(b) *The associated canonical height function extends $\mathbb{R}$-linearly to a positive definite quadratic form*

$$\hat{h}_{A,D} : A(\bar{k}) \otimes \mathbb{R} \longrightarrow \mathbb{R}.$$

In particular, if $P_1,\dots,P_r \in A(\bar{k}) \otimes \mathbb{R}$ are linearly independent, then the height regulator

$$\det\bigl(\langle P_i,P_j\rangle_D\bigr)_{1\le i,j\le r}$$

is strictly greater than 0.

Remark. As pointed out by Cassels, the fact that $\hat{h}_{A,D}$ is a positive definite quadratic form on $A(\bar{k})/(\text{torsion})$ does not, by itself, imply that $\hat{h}_{A,D}$ is positive definite on $A(\bar{k})\otimes\mathbb{R}$. For example, consider the group $\Lambda = \mathbb{Z}+\mathbb{Z}\sqrt{2}$ as a subgroup of $\mathbb{R}$ and the quadratic form on Λ defined by

$$q : \Lambda \to \mathbb{R}, \qquad q(x) = |x|^2.$$

(That is, $q\left(a+b\sqrt{2}\right) = a^2+2b^2+2ab\sqrt{2}$.) It is clear that q is positive definite on Λ, since $\sqrt{2}$ is irrational. However, it is equally clear that q is not positive definite on $\Lambda\otimes\mathbb{R}\cong\mathbb{R}\oplus\mathbb{R}$. For example, $q\left(a+b\sqrt{2}\right)=0$ for $(a,b)=\left(\sqrt{2},-1\right)$. A closer analysis shows that the problem occurs because the set of values $q(\Lambda)$ is not a discrete subset of $\mathbb{R}$; see Corollary B.5.4.1 below.

PROOF. (a) Let $P\in A(\bar{k})$. Proposition B.4.2(a) says that $\hat{h}_{A,D}(P)\geq 0$, with equality if and only if P is preperiodic for (say) the multiplication-by-2 map. It is clear that such preperiodic points are torsion points. Conversely, if P is a point of order n, then using Euler's formula $2^{\phi(n)}\equiv 1 \pmod{n}$, we see that $[2^{\phi(n)}]P=P$, so P is preperiodic.
(b) Let $P\in A(\bar{k})\otimes\mathbb{R}$ be a point with $\hat{h}_{A,D}(P)=0$. We can write P as a linear combination

$$P=a_1P_1+a_2P_2+\cdots+a_rP_r$$

with $a_i\in\mathbb{R}$ and $P_i\in A(\bar{k})$. Replacing k by a finite extension if necessary, we may assume that each P_i is in $A(k)$.

Let $V=\mathbb{R}P_1+\mathbb{R}P_2+\cdots+\mathbb{R}P_r$ be the span of the P_i's in $A(\bar{k})\otimes\mathbb{R}$, and let Λ be the image of $\mathbb{Z}P_1+\mathbb{Z}P_2+\cdots+\mathbb{Z}P_r$ in V. Thus V is a finite-dimensional real vector space, Λ is a finitely generated subgroup of V, and $V=\Lambda\otimes\mathbb{R}$, so Λ is a lattice in V. Further, (B.5.1(d)) tells us that the canonical height $\hat{h}_{A,D}$ induces a quadratic form on V. We will denote this quadratic form by $q:V\to\mathbb{R}$.

Our first observation is that q is positive definite on the lattice Λ. This follows from (a), since the kernel of $A(\bar{k})\to A(\bar{k})\otimes\mathbb{R}$ is precisely the torsion subgroup of $A(\bar{k})$. Thus $\hat{h}_{A,D}$ induces a positive definite quadratic form on the image of $A(\bar{k})$ in $A(\bar{k})\otimes\mathbb{R}$. In particular, it is positive definite on Λ.

We will next verify that for any constant B, the set

$$\{\lambda\in\Lambda \mid q(\lambda)\leq B\}$$

is finite. To see this, let λ be the image of some point $Q\in A(k)$. Then

$$q(\lambda)\leq B \Longrightarrow \hat{h}_{A,D}(Q)\leq B \Longrightarrow h_{A,D}(Q)\leq B+C,$$

since $\hat{h}_{A,D}$ and $h_{A,D}$ differ by a bounded amount. Now we use the assumption that D is ample to apply (B.3.2(g)) and conclude that

$$\{Q\in A(k) \mid h_{A,D}(Q)\leq B+C\}$$

is finite.

We now know that q is a positive definite quadratic form on the lattice Λ and that there are only finitely many elements $\lambda\in\Lambda$ with bounded

$q(\lambda)$. The fact that q remains positive definite when extended to $\Lambda \otimes \mathbb{R}$ is then a simple corollary of Minkowski's theorem on lattice points in symmetric domains. For completeness, we give the proof of Minkowski's theorem, followed by the application to quadratic forms, thereby completing the proof of Theorem B.5.3. □

Theorem B.5.4. (Minkowski) *Let Λ be a lattice in $\mathbb{R}^r$, let F be a fundamental domain for $\mathbb{R}^r/\Lambda$, and let $U \subset \mathbb{R}^r$ be a symmetric convex set. (That is, $x \in U$ implies $-x \in U$, and $x, y \in U$ implies $tx + (1-t)y \in U$ for all $0 \le t \le 1$.) If $\operatorname{vol}(U) > 2^r \operatorname{vol}(F)$, then $U \cap \Lambda$ contains a nonzero vector.*

PROOF. Let $W = \frac{1}{2}U = \{\frac{1}{2}x \mid x \in U\}$. Suppose that

$$(W+\lambda) \cap (W+\mu) = \emptyset \quad \text{for all } \lambda, \mu \in \Lambda \text{ with } \lambda \ne \mu. \tag{$*$}$$

Then

$$\begin{aligned}\frac{1}{2^r}\operatorname{vol}(U) &= \operatorname{vol}(W)\\ &= \operatorname{vol}\Bigl(W \cap \bigcup_{\lambda\in\Lambda}(F+\lambda)\Bigr)\\ &= \sum_{\lambda\in\Lambda}\operatorname{vol}\bigl(W \cap (F+\lambda)\bigr)\\ &= \sum_{\lambda\in\Lambda}\operatorname{vol}\bigl((W-\lambda)\cap F\bigr)\\ &= \operatorname{vol}\Bigl(\bigcup_{\lambda\in\Lambda}(W-\lambda)\cap F\Bigr) \quad \text{from } (*),\\ &\le \operatorname{vol}(F).\end{aligned}$$

This contradicts the assumption that $\operatorname{vol}(U) > 2^r \operatorname{vol}(F)$, so we conclude that $(*)$ is false. Hence we can find distinct elements $\lambda, \mu \in \Lambda$ and points $x, y \in W$ such that $x + \lambda = y + \mu$. Then the symmetry and convexity of U allows us to compute

$$0 \ne \lambda - \mu = y - x \in W + W = \frac{1}{2}U + \frac{1}{2}U \subset U,$$

so $U \cap \Lambda$ contains the nonzero vector $\lambda - \mu$. □

Corollary B.5.4.1. *Let Λ be a free abelian group of finite rank. Let $q : \Lambda \to \mathbb{R}$ be a quadratic form with the following two properties:*
(i) *For all $\lambda \in \Lambda$, $q(\lambda) \ge 0$, with equality if and only if $\lambda = 0$.*
(ii) *For all B, the set $\bigl\{\lambda \in \Lambda \mid q(\lambda) \le B\bigr\}$ is finite.*
Then q extends to a positive definite quadratic from on $\Lambda \otimes \mathbb{R}$.

PROOF. Let $V = \Lambda\otimes\mathbb{R}$, and by abuse of notation let $q : V \to \mathbb{R}$ also denote the extension of $q : \Lambda \to \mathbb{R}$ to V. A standard result from linear algebra

says that a real quadratic form on a real vector space can be diagonalized with 0's, 1's, and −1's on the diagonal. In other words, we can choose an isomorphism $V \cong \mathbb{R}^r$ such that if $\mathbf{x} = (x_1, \ldots, x_r) \in V$, then

$$q(\mathbf{x}) = x_1^2 + x_2^2 + \cdots + x_s^2 - x_{s+1}^2 - x_{s+2}^2 - \cdots - x_{s+t}^2.$$

See, for example, Lang [2, Chapter 7, Sections 3,7] or van der Waerden [1, Section 12.7]. The integers s and t are uniquely determined by q and satisfy $s + t \le r = \dim(V)$. Notice that q is positive definite on V if and only if $s = r$.

Let

$$L = \inf\{q(\lambda) \mid \lambda \in \Lambda,\ \lambda \neq 0\}.$$

The given properties (i) and (ii) of q ensure that $L > 0$. Now for each pair of positive numbers $\delta, \varepsilon > 0$ we consider the set

$$U(\delta,\varepsilon) = \left\{ \mathbf{x} = (x_1, \ldots, x_r) \in \mathbb{R}^r \;\middle|\; \sum_{i=1}^{s} x_i^2 \le \delta \quad \text{and} \quad \sum_{i=1}^{t} x_{s+i}^2 \le \varepsilon \right\}.$$

The set $U(\delta,\varepsilon)$ is clearly convex and symmetric about the origin. Further, the definition of L ensures that $U(L/2,\varepsilon) \cap \Lambda = \{0\}$ for all $\varepsilon > 0$. It follows from Minkowski's theorem (B.5.4) that $\operatorname{vol}\big(U(L/2,\varepsilon)\big)$ is bounded as $\varepsilon \to \infty$.

On the other hand, it is clear from the definition of $U(\delta,\varepsilon)$ that the volume of $U(L/2,\varepsilon)$ is infinite if $s + t < r$, and that the volume grows like $\varepsilon^{t/2}$ as $\varepsilon \to \infty$ if $s + t = r$. Hence the boundedness of $\operatorname{vol}\big(U(L/2,\varepsilon)\big)$ implies that $s = r$, which completes the proof that q is positive definite on V. □

We have now developed in some detail the theory of canonical heights associated to symmetric divisor classes on an abelian variety. There is an analogous theory for antisymmetric divisor classes, that is, divisors that satisfy $[-1]^*D \sim -D$. The associated canonical heights turn out to be linear rather than quadratic.

Theorem B.5.5. *Let A/k be an abelian variety defined over a number field, and let $D \in \operatorname{Div}(A)$ be a divisor whose divisor class is antisymmetric (i.e., $[-1]^*D \sim -D$). There is a canonical height function*

$$\hat{h}_{A,D} : A(\bar{k}) \longrightarrow \mathbb{R},$$

with the following properties:

(a) $\hat{h}_{A,D}(P) = h_{A,D}(P) + O(1)$ *for all* $P \in A(\bar{k})$.

(b) $\hat{h}_{A,D}(P + Q) = \hat{h}_{A,D}(P) + \hat{h}_{A,D}(Q)$ *for all* $P, Q \in A(\bar{k})$.

Thus $\hat{h}_{A,D}$ is a homomorphism from $A(\bar{k})$ to $\mathbb{R}$. Further, the function $\hat{h}_{A,D}$ is uniquely determined by properties (a) and (b).

PROOF. The proof is almost the same as the proof of (B.5.1), so we merely give a sketch. Corollary A.7.2.5 says that $[m]^*D \sim mD$, so (B.4.1) says that there is a canonical height $\hat{h}_{A,D}$ with respect to (say) the map [2]. From (B.4.1), this height satisfies (a), and it can be computed as the limit

$$\hat{h}_{A,D}(P) = \lim_{n\to\infty} 2^{-n} h_{A,D}\big([2^n]P\big).$$

To prove (b), we use the relation

$$h_{A,D}(P+Q) = h_{A,D}(P) + h_{A,D}(Q) + O(1) \qquad \text{for all } P, Q \in A(\bar{k})$$

from (B.3.4(c)). Now replace P and Q by 2^nP and 2^nQ, divide by 2^n, and let $n \to \infty$ to obtain (b). Finally, suppose that $\hat{h}$ and $\hat{h}'$ both satisfy (a) and (b). Then the difference $\hat{h} - \hat{h}'$ is a bounded homomorphism $A(\bar{k}) \to \mathbb{R}$. The image of this homomorphism is a bounded subgroup of $\mathbb{R}$. But $\mathbb{R}$ has no bounded subgroups other than $\{0\}$, so $\hat{h} = \hat{h}'$. This proves the uniqueness and completes the proof of the theorem. □

An arbitrary divisor can (almost) be written as the sum of a symmetric and an antisymmetric divisor. So combining our Theorems B.5.1 and B.5.5, we will be able to construct canonical heights for all divisors. In order to state our result, it is helpful to recall the following definition.

Definition. Let A and B be abelian groups, and assume that B is uniquely 2-divisible. (That is, assume that the doubling map $B \to B$, $b \mapsto 2b$, is an isomorphism. For example, $\mathbb{R}$ is uniquely 2-divisible.) A *quadratic function* on A with values in B is a function $h : A \to B$ satisfying

$$h(P+Q+R)-h(P+Q)-h(P+R)-h(Q+R)+h(P)+h(Q)+h(R)-h(0) = 0 \qquad \text{for all } P, Q, R \in A.$$

Equivalently, $h : A \to B$ is a quadratic function if the associated pairing

$$\langle\cdot,\cdot\rangle_h : A \times A \longrightarrow B, \qquad \langle P, Q\rangle_h = \frac{h(P+Q) - h(P) - h(Q) + h(0)}{2},$$

is bilinear. We note that a quadratic function can be written uniquely as the sum of a quadratic form, a linear form, and a constant; see Exercise B.9.

Theorem B.5.6. *Let A/k be an abelian variety defined over a number field, and let $D \in \mathrm{Div}(A)$ be a divisor on A.*
(a) *There is a unique quadratic function*

$$\hat{h}_{A,D} : A(\bar{k}) \longrightarrow \mathbb{R}$$

satisfying $\hat{h}_{A,D} = h_{A,D} + O(1)$ *and* $\hat{h}_{A,D}(0) = 0$. *It is called the canonical height on* A *relative to* D.
(b) *The canonical height* $\hat{h}_{A,D}$ *depends only on the divisor class of* D.
(c) *Let* $D, E \in \operatorname{Div}(A)$. *Then* $\hat{h}_{A,D+E} = \hat{h}_{A,D} + \hat{h}_{A,E}$.
(d) *Let* B/k *be another abelian variety, and let* $\phi : B \to A$ *be a morphism. Then*

$$\hat{h}_{B,\phi^* D} = \hat{h}_{A,D} \circ \phi - \hat{h}_{A,D}(\phi(0)).$$

(e) *There is a unique quadratic form* $\hat{q}_{A,D} : A(\bar{k}) \to \mathbb{R}$ *and a unique linear form* $\hat{\ell}_{A,D} : A(\bar{k}) \to \mathbb{R}$ *such that*

$$\hat{h}_{A,D} = \hat{q}_{A,D} + \hat{\ell}_{A,D}.$$

More precisely,

$$\hat{q}_{A,D} = \frac{1}{2}\hat{h}_{A,D+[-1]^*D} \qquad \text{and} \qquad \hat{\ell}_{A,D} = \frac{1}{2}\hat{h}_{A,D-[-1]^*D}.$$

Hence if D *represents a symmetric divisor class* ($[-1]^*D \sim D$), *then* $\hat{h}_{A,D} = \hat{q}_{A,D}$; *and if* D *represents an antisymmetric divisor class* ($[-1]^*D \sim -D$), *then* $\hat{h}_{A,D} = \hat{\ell}_{A,D}$.

PROOF. Define divisors $D^+ = D+[-1]^*D$ and $D^- = D-[-1]^*D$. It is clear that D^+ is symmetric and D^- is antisymmetric. Applying Theorem B.5.1 to the divisor D^+, we obtain a quadratic form $\hat{h}_{A,D^+}$ on $A(\bar{k})$; and applying Theorem B.5.5 to the divisor D^-, we obtain a linear form $\hat{h}_{A,D^-}$ on $A(\bar{k})$. Define

$$\hat{h}_{A,D} : A(\bar{k}) \longrightarrow \mathbb{R}, \qquad \hat{h}_{A,D} = \frac{1}{2}\left(\hat{h}_{A,D^+} + \hat{h}_{A,D^-}\right).$$

Then $\hat{h}_{A,D}$ is certainly a quadratic function with $\hat{h}_{A,D}(0) = 0$, since it is a sum of a linear form and a quadratic form. Further,

$$\begin{aligned}
2\hat{h}_{A,D} &= \hat{h}_{A,D^+} + \hat{h}_{A,D^-} && \text{definition of } \hat{h}_{A,D} \\
&= h_{A,D^+} + h_{A,D^-} + O(1) && \text{from (B.5.1) and (B.5.5)} \\
&= h_{A,D+[-1]^*D} + h_{A,D-[-1]^*D} + O(1) && \text{definition of } D^+ \text{ and } D^- \\
&= 2h_{A,D} + O(1) && \text{additivity (B.3.2(c))}.
\end{aligned}$$

This shows that $\hat{h}_{A,D}$ has the desired properties, which proves the existence part of (a).

Now suppose that $\hat{h}'_{A,D}$ is another function satisfying the properties in (a). Let $f = \hat{h}_{A,D} - \hat{h}'_{A,D}$. Then f is a bounded quadratic function $f : A(\bar{k}) \to \mathbb{R}$ and satisfies $f(0) = 0$. In particular, the bilinear form

$$\langle \cdot, \cdot \rangle_f : A(\bar{k}) \times A(\bar{k}) \to \mathbb{R}, \quad \langle P, Q \rangle_f = \frac{f(P+Q) - f(P) - f(Q) + f(0)}{2},$$

associated to f is bounded. But $\langle nP, nQ\rangle_f = n^2\langle P, Q\rangle_f$, so the bilinear form is identically 0. Thus $f(P+Q) + f(0) = f(P) + f(Q)$. But this implies that $f(nP) - f(0) = n\big(f(P) - f(0)\big)$, so the boundedness of f implies that $f(P) = f(0)$ for all P. Since $f(0) = 0$, this proves that $\hat{h}_{A,D} = \hat{h}'_{A,D}$, which completes the proof of uniqueness.

(b) If $D' \sim D$, then $h_{A,D} = h_{A,D'} + O(1)$ from (B.3.2(d)). Thus $\hat{h}_{A,D'}$ satisfies the conditions in (a) for the divisor D, so by the uniqueness assertion in (a), we have $\hat{h}_{A,D'} = \hat{h}_{A,D}$.

(c) We have $h_{A,D+E} = h_{A,D} + h_{A,E} + O(1)$ from (B.3.2(c)). Thus the function $\hat{h}_{A,D} + \hat{h}_{A,E}$ satisfies the conditions in (a) for the divisor $D+E$, so by the uniqueness assertion in (a), we have $\hat{h}_{A,D} + \hat{h}_{A,E} = \hat{h}_{A,D+E}$.

(d) To ease notation, let $f = \hat{h}_{A,D} \circ \phi - \hat{h}_{A,D}\big(\phi(0)\big)$. Using (B.3.2(b)) and properties of canonical heights, we see that

$$f = h_{A,D} \circ \phi + O(1) = h_{A,\phi^* D} + O(1).$$

Next, it is clear that $f(0) = 0$. Finally, every morphism between abelian varieties is the composition of a homomorphism and a translation (A.7.1.2), so the fact that $\hat{h}_{A,D}$ is a quadratic function implies the same for f. Thus f has all of the properties to be $\hat{h}_{A,\phi^* D}$, so by the uniqueness assertion in A, it is equal to $\hat{h}_{A,\phi^* D}$.

(e) In (a) we have already written $\hat{h}_{A,D}$ as the sum of the quadratic form $\frac{1}{2}\hat{h}_{A,D^+}$ and the linear form $\frac{1}{2}\hat{h}_{A,D^-}$. This gives the existence. We will leave the proof of uniqueness to the reader. (More generally, any quadratic function has a unique decomposition as a sum of a quadratic form, a linear form, and a constant; see Exercise B.9.) The final statements in (e) are then clear, since if D is symmetric, then $D^+ \sim 2D$ and $D^- \sim 0$, while if D is antisymmetric, then $D^+ \sim 0$ and $D^- \sim 2D$. □

Remark B.5.7. The first three parts of Theorem B.5.6 combine to say that there is a homomorphism

$$\operatorname{Pic}(A) \longrightarrow \{\text{quadratic functions } f : A(\bar{k}) \to \mathbb{R}\}$$

that sends a divisor class c to the canonical height $\hat{h}_{A,c}$. Here $\hat{h}_{A,c}$ is defined to be $\hat{h}_{A,D}$ for any divisor D in the divisor class c. It is well-defined independent of the choice of D by (B.5.6(b)).

We recall from Proposition A.7.3.2 that on an abelian variety, the group of antisymmetric divisor classes is equal to the group $\operatorname{Pic}^0(A)$ of divisor classes that are algebraically equivalent to zero. Theorem B.5.6 implies that if $c \in \operatorname{Pic}^0(A)$, then the corresponding canonical height $\hat{h}_{A,c}$ is a homomorphism $A(\bar{k}) \to \mathbb{R}$. We also recall that $\operatorname{Pic}^0(A)$ is naturally isomorphic to an abelian variety $\hat{A}$, called the dual abelian variety of A, and that there is a Poincaré divisor class $\mathcal{P}$ on $A \times \hat{A}$ determined by certain functorial properties. (See Section A.7.3, especially Theorem A.7.3.4). The following theorem expands on these observations.

Theorem B.5.8. *Define a canonical height pairing by the formula*

$$[\,\cdot\,,\,\cdot\,]_A : A(\bar{k}) \times \mathrm{Pic}^0(A) \longrightarrow \mathbb{R}, \qquad [P,c]_A = \hat{h}_{A,c}(P).$$

(a) *The canonical height pairing* $[\,\cdot\,,\,\cdot\,]_A$ *is bilinear, and its kernel on either side consists of the elements of finite order.*
(b) *Let* $\hat{A}$ *be the dual abelian variety to* A*, and let* $\mathcal{P} \in A \times \hat{A}$ *be the Poincaré divisor class. Then with the natural identification of* $\hat{A}$ *and* $\mathrm{Pic}^0(A)$,

$$[P,c]_A = \hat{h}_{A\times\hat{A},\mathcal{P}}(P,c) \qquad \text{for all } (P,c) \in A \times \hat{A}.$$

Proof. The pairing is well-defined from (B.5.6(a,b)). Divisor classes in $\mathrm{Pic}^0(A)$ are antisymmetric from (A.7.3.2), so (B.5.6(e)) implies that $[P,c]_A$ is linear in P. Similarly, the linearity in c follows from (B.5.6(c)). This proves that $[\,\cdot\,,\,\cdot\,]_A$ is a well-defined bilinear pairing.

For any $Q \in A$, let $t_Q : A \to A$ be the translation-by-Q map, $t_Q(P) = P + Q$. We also fix a symmetric ample divisor class $\xi \in \mathrm{Pic}(A)$, and we recall from Theorem A.7.3.1 that the map

$$A \longrightarrow \mathrm{Pic}^0(A), \qquad Q \longmapsto t_Q^*\xi - \xi,$$

is surjective.

Now fix a point $P \in A(\bar{k})$, and suppose that $[P,c]_A = 0$ for all $c \in \mathrm{Pic}^0(A)$. Then $[P, t_Q^*\xi - \xi]_A = 0$ for all $Q \in A(\bar{k})$. Using the definition of the pairing and standard properties of height functions, we find that

$$\begin{aligned} 0 &= [P, t_Q^*\xi - \xi]_A \\ &= \hat{h}_{A,t_Q^*\xi-\xi}(P) \quad (\text{definition of } [\,\cdot\,,\,\cdot\,]_A) \\ &= \hat{h}_{A,t_Q^*\xi}(P) - \hat{h}_{A,\xi}(P) \quad (\text{linearity (B.5.6(c))}) \\ &= \hat{h}_{A,\xi}\big(t_Q(P)\big) - \hat{h}_{A,\xi}(Q) - \hat{h}_{A,\xi}(P) \quad (\text{from (B.5.6(d))}) \\ &= \hat{h}_{A,\xi}(P+Q) - \hat{h}_{A,\xi}(Q) - \hat{h}_{A,\xi}(P) \quad (\text{definition of } t_Q). \end{aligned}$$

This holds for all $Q \in A(\bar{k})$. In particular, we may take $Q = P$. Now ξ is symmetric by assumption, so (B.5.6(e)) says that $\hat{h}_{A,\xi}(2P) = 4\hat{h}_{A,\xi}(P)$. We conclude that $\hat{h}_{A,\xi}(P) = 0$. Now the assumption that ξ is ample and (B.5.3(a)) imply that P is a torsion point.

Similarly, fix a divisor class $c \in \mathrm{Pic}^0(A)$, and suppose that $[P,c]_A = 0$ for all $P \in A(\bar{k})$. We know that c can be written in the form $c = t_Q^*\xi - \xi$ for some $Q \in A(\bar{k})$, and just as above we compute

$$0 = [P,c]_A = [P, t_Q^*\xi - \xi]_A = \hat{h}_{A,\xi}(P+Q) - \hat{h}_{A,\xi}(Q) - \hat{h}_{A,\xi}(P).$$

This time we set $P = Q$ to get $\hat{h}_{A,\xi}(Q) = 0$, so Q is a torsion point. But the map $Q \mapsto t_Q^*\xi - \xi$ from A to $\operatorname{Pic}^0(A)$ is a homomorphism (A.7.2.9), so $c = t_Q^*\xi - \xi$ has finite order in $\operatorname{Pic}^0(A)$.

(b) We identify the dual abelian variety $\hat{A}$ of A with $\operatorname{Pic}^0(A)$, and for each $c \in \hat{A}$, we define an inclusion

$$i_c : A \longrightarrow A \times \hat{A}, \qquad P \longmapsto (P, c).$$

By the definition of the dual abelian variety and its Poincaré divisor class $\mathcal{P}$ (see Section A.7.3), we have $i_c^*\mathcal{P} = c$. It follows from standard properties of height functions, specifically (B.5.6(d)), that

$$\begin{aligned}[P,c]_A &= \hat{h}_{A,c}(P) = \hat{h}_{A,i_c^*\mathcal{P}}(P) = \hat{h}_{A\times\hat{A},\mathcal{P}}\big(i_c(P)\big) - \hat{h}_{A\times\hat{A},\mathcal{P}}\big(i_c(0)\big) \\ &= \hat{h}_{A\times\hat{A},\mathcal{P}}(P,c) - \hat{h}_{A\times\hat{A},\mathcal{P}}(0,c).\end{aligned}$$

Now consider the inclusion $\hat{\imath}_0 : \hat{A} \to A \times \hat{A}$ defined by $\hat{\imath}_0(c) = (0, c)$. The definition of the Poincaré divisor class implies that $\hat{\imath}_0^*\mathcal{P} = 0$, so

$$\begin{aligned}0 = \hat{h}_{\hat{A},\hat{\imath}_0^*\mathcal{P}}(c) &= \hat{h}_{A\times\hat{A},\mathcal{P}}\big(\hat{\imath}_0(c)\big) - \hat{h}_{A\times\hat{A},\mathcal{P}}\big(\hat{\imath}_0(0)\big) \\ &= \hat{h}_{A\times\hat{A},\mathcal{P}}(0,c) - \hat{h}_{A\times\hat{A},\mathcal{P}}(0,0) = \hat{h}_{A\times\hat{A},\mathcal{P}}(0,c).\end{aligned}$$

Combining this with the formula for $[P, c]_A$ proven above gives the desired result. □

We now apply the theory of canonical height and functorial properties of Picard and Albanese varieties to strengthen the algebraic equivalence property of Weil's Height Machine (Theorem B.3.2(f)).

Theorem B.5.9. *Let k be a number field and V/k a smooth projective variety. Let $D, E \in \operatorname{Div}(V)$ be divisors with D ample and E algebraically equivalent to 0. Then there is a constant $c > 0$ such that*

$$h_{V,E}(P) \le c\sqrt{h_{V,D}(P) + 1} \quad \text{for all } P \in V(\bar{k}).$$

PROOF. Let $A = \operatorname{Alb}(V)$ be the Albanese variety of V, and let $\pi : V \to A$ be the universal map from V to A. We recall from Proposition A.7.3.6 that there exists a divisor $E_1 \in \operatorname{Div}(A)$, algebraically equivalent to 0 on A, such that $E = \pi^*E_1$.

Let $D_1 \in \operatorname{Div}(A)$ be any symmetric ample divisor on A. Since E_1 is algebraically equivalent to 0, Theorem A.7.3.1(c) tells us that there is a point $a \in A(\bar{k})$ such that $E_1 \sim t_a^*D_1 - D_1$, where $t_a : A \to A$ is the translation-by-a map.

Now let $P \in V(\bar{k})$, and for notational convenience, let $Q = \pi(P)$. We compute

$$\begin{aligned}
h_{V,E}(P) &= h_{V,\pi^* E_1}(P) + O(1) && \text{since } E = \pi^* E_1 \\
&= h_{A,E_1}(Q) + O(1) && \text{by functoriality (B.3.2(b))} \\
&= \hat{h}_{A,E_1}(Q) + O(1) && \text{by (B.5.5(a))} \\
&= \hat{h}_{A,t_a^* D_1}(Q) - \hat{h}_{A,D_1}(Q) + O(1) && \text{by (B.5.6(c))} \\
&= \hat{h}_{A,D_1}(t_a(Q)) - \hat{h}_{A,D_1}(t_a(O)) - \hat{h}_{A,D_1}(Q) + O(1) \\
& && \text{by (B.5.6(d)) applied to } t_a \\
&= \hat{h}_{A,D_1}(Q+a) - \hat{h}_{A,D_1}(a) - \hat{h}_{A,D_1}(Q) + O(1) \\
&= 2\langle Q, a\rangle_{D_1} + O(1) && \text{by definition (B.5.1(d))} \\
&\le 2\sqrt{\hat{h}_{A,D_1}(Q)\hat{h}_{A,D_1}(a)} + O(1) && \text{by Cauchy–Schwarz.}
\end{aligned}$$

Note that the last inequality follows by applying the Cauchy–Schwarz inequality to the positive definite quadratic form $\hat{h}_{A,D_1}$ and its associated bilinear form $\langle\cdot,\cdot\rangle_{D_1}$.

Similarly, we have

$$\begin{aligned}
\hat{h}_{A,D_1}(Q) &= h_{A,D_1}(\pi(P)) + O(1) && \text{by (B.5.1(a)) and } Q = \pi(P) \\
&= h_{V,\pi^* D_1}(P) + O(1) && \text{by functoriality (B.3.2(b))} \\
&\le c h_{V,D}(P) + O(1) && \text{for some } c > 0, \text{ since } D \text{ is ample.}
\end{aligned}$$

Substituting this estimate into the previous one and adjusting the constants gives the desired result. □

B.6. Counting Rational Points on Varieties

The coarsest measure of the set of rational points on an algebraic variety is whether the set is finite or infinite. Finiteness theorems are generally proven by showing that the set of rational points is a set of bounded height. When there are infinitely many rational points, heights can be used to define a counting function whose asymptotic behavior often encodes deep arithmetic information.

Definition. Let V/k be a projective variety defined over a number field k. Fix a multiplicative height function H_V on V relative to some ample divisor D. The *counting function of* $V(k)$ is

$$N\big(V(k),T\big) = \#\{P \in V(k) \mid H_V(P) \le T\}.$$

If we need to specify the divisor D or even the particular height function H_V, we will use the notation $N\big(V(k),D,T\big)$ or $N\big(V(k),H_V,T\big)$. Similarly, if U is an open subset of V, we define a counting function $N\big(U(k),T\big)$ for $U(k)$ by only counting points in U.

Generally, the goal is to describe the behavior of $N\big(V(k),T\big)$ as $T \to \infty$ in terms of geometric invariants of the variety V and arithmetic invariants of the field k. This goal has been most fully realized in the case of curves, as described in the following result.

Theorem B.6.1. *Let k be a number field, let C/k be a smooth curve of genus g, and assume that $C(k)$ is not empty. Then there are constants a and b, which depend on C/k and on the height used in the counting function, such that*

$$N\big(C(k),T\big) \sim \begin{cases} aT^b & \text{if } g=0 \ (a,b>0), \\ a(\log T)^b & \text{if } g=1 \ (a>0,\ b\ge 0), \\ a & \text{if } g\ge 2. \end{cases}$$

PROOF. If $g=0$, then $C(k)\cong \mathbb{P}^1(k)$. (Note we are assuming that $C(k)\neq \emptyset$.) Later in this section we will give a much more precise description of the counting function on projective space; see (B.6.2) below. Similarly, if $g=1$, then C is an abelian variety, and we will describe the counting function on abelian varieties below (B.6.3). Finally, the assertion for genus $g\ge 2$ is that $C(k)$ is finite. This is Faltings' theorem (originally Mordell's conjecture), which we will prove in Part E. As a warm-up, in this section we will prove a weaker result of Mumford saying that the counting function of a curve of genus $g\ge 2$ satisfies $N\big(C(k),T\big) \ll \log\log T$; see (B.6.5). □

Remark. The theorem is actually still valid for singular curves, since they will have the same number of points as their normalization, up to $O(1)$.

It is possible to give a very precise description of the counting function on projective space. For $k=\mathbb{Q}$, this result is classical. The general case, which is more complicated, is due to Schanuel.

Theorem B.6.2. (Schanuel [1]) *Let k be a number field of degree d, let $n\ge 1$ be an integer, and let $N\big(\mathbb{P}^n(k),T\big)$ be the counting function on $\mathbb{P}^n$ relative to the usual multiplicative height H_k. (Note that this is the height relative to k, not the absolute height.) Then there is a constant $a(k,n)>0$ such that*

$$N\big(\mathbb{P}^n(k),T\big) = a(k,n)T^{n+1} + \begin{cases} O(T\log T) & \text{if } k=\mathbb{Q} \text{ and } n=1, \\ O(T^{n+1-1/d}) & \text{otherwise.} \end{cases}$$

More precisely, the constant $a(k,n)$ is equal to

$$a(k,n) = \frac{hR/w}{\zeta_k(n+1)}\left(\frac{2^{r_1}(2\pi)^{r_2}}{\sqrt{D_k}}\right)^{n+1}(n+1)^{r_1+r_2-1},$$

where

$$\begin{aligned} h &= \textit{class number of } k, \\ R &= \textit{regulator of } k, \\ w &= \textit{number of roots of unity in } k, \\ \zeta_k &= \textit{zeta function of } k, \\ r_1 &= \textit{number of real embeddings of } k, \\ r_2 &= \textit{number of complex embeddings of } k, \\ D_k &= \textit{absolute value of the discriminant of } k/\mathbb{Q}. \end{aligned}$$

PROOF. We will illustrate the proof by doing the case $k = \mathbb{Q}$. See Schanuel [1] or Lang [6, Chapter 3, Theorem 5.3] for the general case. We normalize the homogeneous coordinates of points in $P \in \mathbb{P}^n(\mathbb{Q})$ by writing

$$P = (x_0, \dots, x_n) \text{ with } x_0, \dots, x_n \in \mathbb{Z} \text{ and } \gcd(x_0, \dots, x_n) = 1.$$

This determines the coordinates up to multiplication by ± 1, so we will need to divide our final count by 2. Note that with this normalization, $H(P) = \max |x_i|$.

For any vector $\mathbf{x} = (x_0, \dots, x_n) \in \mathbb{A}^{n+1}(\mathbb{Z})$, let $|\mathbf{x}| = \max |x_i|$ and $\gcd(\mathbf{x}) = \gcd(x_i)$. For integers $d \geq 1$, we define two counting functions,

$$\begin{aligned} M(T) &= \#\{\mathbf{x} \in \mathbb{A}^{n+1}(\mathbb{Z}) \mid \mathbf{x} \neq 0 \text{ and } |\mathbf{x}| \leq T\}, \\ M^*(T, d) &= \#\{\mathbf{x} \in \mathbb{A}^{n+1}(\mathbb{Z}) \mid \gcd(\mathbf{x}) = d \text{ and } |\mathbf{x}| \leq T\}. \end{aligned}$$

From our discussion above, we have $N(\mathbb{P}^n(\mathbb{Q}), T) = \frac{1}{2} M^*(T, 1)$.

We observe that if $\mathbf{x} = (x_0, \dots, x_n)$ is counted in $M^*(T, d)$, then $\mathbf{x}/d = (x_0/d, \dots, x_n/d)$ will be counted in $M^*(T/d, 1)$; and conversely, if $\mathbf{x}$ is counted in $M^*(T/d, 1)$, then $d\mathbf{x}$ will be counted in $M^*(T, d)$. This gives the useful relation $M^*(T, d) = M^*(T/d, 1)$. We also note that every point $\mathbf{x}$ counted in $M(T)$ has $\gcd(\mathbf{x}) \geq 1$, so $\mathbf{x}$ is counted in exactly one $M^*(T, d)$. Combining these two remarks gives

$$M(T) = \sum_{d \geq 1} M^*(T, d) = \sum_{d \geq 1} M^*(T/d, 1).$$

Note that the sum is finite, since $M^*(T/d, 1) = 0$ if $d > T$. Applying Möbius inversion, we find that

$$M^*(T, 1) = \sum_{d \geq 1} \mu(d) M(T/d).$$

(See Apostol [1] for information about the Möbius function μ and Möbius inversion.)

But the counting function $M(T)$ is easy to compute,

$$M(T) = \left(\#\{x \in \mathbb{Z} \mid |x| \leq T\}\right)^{n+1} - 1 = \left(2[T] + 1\right)^{n+1} - 1,$$

where $[T]$ denotes the greatest integer in T. Note that we subtract 1 because the point $(0,0,\dots,0)$ is not counted in $M(T)$. In particular, $M(T)=0$ for $0 \le T < 1$. We compute

$$\begin{aligned}
M^*(T,1) &= \sum_{d\ge 1} \mu(d) M(T/d) \quad \text{(from above)} \\
&= \sum_{1\le d\le T} \mu(d)\bigl((2[T/d]+1)^{n+1}-1\bigr) \\
&= \sum_{1\le d\le T} \mu(d)\bigl(2T/d+O(1)\bigr)^{n+1} \\
&= \sum_{1\le d\le T} \mu(d)\bigl((2T/d)^{n+1}+O(T/d)^n\bigr) \\
&= (2T)^{n+1}\sum_{d\ge 1}\mu(d)d^{-n-1} \\
&\qquad - (2T)^{n+1}\sum_{d>T}\mu(d)d^{-n-1} + O\biggl(T^n \sum_{1\le d\le T} d^{-n}\biggr).
\end{aligned}$$

The first sum is the main term, since $\sum_{d\ge 1}\mu(d)d^{-n-1}$ is equal to $1/\zeta(n+1)$. The second sum contributes to the error, since $\sum_{d>T} d^{-n-1} = O(T^{-n})$. Finally, the third sum (in the big-O) also gives an error term, since

$$T^n \sum_{1\le d\le T} d^{-n} = \begin{cases} O(T^n) & \text{if } n\ge 2, \\ O(T\log T) & \text{if } n=1. \end{cases}$$

Hence

$$N\bigl(\mathbb{P}^n(\mathbb{Q}),T\bigr) = \frac{1}{2}M^*(T,1) = \frac{2^n}{\zeta(n+1)}T^{n+1} + O(T^n) \quad (\text{or } O(T\log T)).$$

This is exactly as stated in Theorem B.6.2, since for $k=\mathbb{Q}$ we have $h=1$, $R=1$, $w=2$, $r_1=1$, $r_2=0$, and $D_{\mathbb{Q}}=1$, so $a(\mathbb{Q},n)=2^n/\zeta_{\mathbb{Q}}(n+1)$. □

Next we will give Néron's description of the counting function of an abelian variety. This result was one of the principal motivations for Néron's construction of canonical height functions. As is often the case in mathematics, tools developed to answer one fundamental question frequently find applications in many other contexts.

Theorem B.6.3. (Néron) *Let k be a number field, let A/k be an abelian variety, and let $\Gamma \subset \operatorname{rank} A(k)$ be a finitely generated group of rank r. There is a constant $a>0$, depending on A/k, Γ, and the height used in the counting function, such that*

$$N(\Gamma,T) = a(\log T)^{r/2} + O\bigl((\log T)^{(r-1)/2}\bigr).$$

PROOF. Let $H_{A,D}$ be the height used in the counting function $N(\Gamma, T)$, let $h_{A,D} = \log H_{A,D}$ be the corresponding logarithmic height, and let $\hat{h}_{A,D}$ be the associated canonical height. (See Theorem B.5.6 for basic properties of the canonical height.) We know that $h_{A,D} = \hat{h}_{A,D} + O(1)$, so it will suffice to prove that

$$N(\Gamma, \hat{h}_{A,D}, T) = aT^{r/2} + O(T^{(r-1)/2}).$$

Let $V = \Gamma \otimes \mathbb{R}$, and let Λ be the image of Γ in V. It suffices to count points in Λ, since the torsion subgroup of $A(k)$, and thus of Γ, is finite (B.4.3(b)). The canonical height can be written in the form $\hat{h}_{A,D} = \hat{q} + \hat{\ell}$, where $\hat{q}$ is a positive definite quadratic form on V and ℓ is a linear form on V. (See (B.5.6(e)) for the decomposition and (B.5.3(b)) for the positivity. Note that $2\hat{q}$ is the canonical height on A relative to the ample symmetric divisor $D + [-1]^*D$.) Now, quadratic forms grow more rapidly than linear forms, precisely, $|\ell| = O(q^{1/2})$, so

$$\hat{h}_{A,D}(x) = \hat{q}(x) + O\big(q(x)^{1/2}\big) \qquad \text{for all } x \in V.$$

It thus suffices to use $\hat{q}$ as our counting function, so we need to prove that

$$N(\Gamma, \hat{q}, T) = aT^{r/2} + O(T^{(r-1)/2}).$$

We can now omit all reference to heights and abelian varieties, since the desired result is a consequence of the following elementary counting lemma.

Lemma B.6.4. *Let $V \cong \mathbb{R}^r$ be a real vector space, let $q : V \to \mathbb{R}$ be a positive definite quadratic form on V, and let $\Lambda \subset V$ be a lattice. Then there is a constant $a = a(q, \Lambda) > 0$ such that*

$$\#\big\{\lambda \in \Lambda \,|\, q(\lambda) \le T\big\} = aT^{r/2} + O(T^{(r-1)/2}).$$

PROOF. This result is a special case of a general counting theorem for lattice points in homogeneously expanding domains. We will be content to prove the special case; see Lang [9, Chapter VI, Section 2] for the general result.

Fix a fundamental domain F for Λ. For example, choose a basis $\lambda_1, \ldots, \lambda_r$ for Λ and take

$$F = \big\{t_1\lambda_1 + \cdots + t_r\lambda_r \mid -\tfrac{1}{2} \le t_i < \tfrac{1}{2}\big\}.$$

For any $\lambda \in \Lambda$, we let F_λ denote the translation of F by λ. By definition, V is equal to the disjoint union $\bigcup_{\lambda \in \Lambda} F_\lambda$.

Let $\|x\| = \sqrt{q(x)}$ denote the norm on V associated to q, and define balls (really ellipsoids)

$$B(T) = \{x \in V \mid \|x\| \le T\}.$$

Let μ be the usual measure on $V = \mathbb{R}^r$. Also let $D = \sup_{x \in F} \|x\|$. We claim that

$$B(T-D) \subset \bigcup_{\lambda \in \Lambda,\, \|\lambda\| \le T} F_\lambda \subset B(T+D).$$

First, suppose that $x \in B(T-D)$. Since V is covered by the F_λ's, we can find some $\lambda \in \Lambda$ such that $x \in F_\lambda$. This means there is a $y \in F$ such that $x = y + \lambda$, which allows us to compute

$$\|\lambda\| = \|x - y\| \le \|x\| + \|y\| \le (T-D) + D = T.$$

(Note that we are using the fact that $x \in B(T-D)$ and that $\|y\| \le D$ for every $y \in F$.) This proves that $x \in F_\lambda$ for some $\lambda \in \Lambda$ satisfying $\|\lambda\| \le T$, which gives the left-hand inclusion.

Second, suppose that $x \in F_\lambda$ for some $\lambda \in \Lambda$ satisfying $\|\lambda\| \le T$. Then $x = y + \lambda$ for some $y \in F$, and hence

$$\|x\| = \|y + \lambda\| \le \|y\| + \|\lambda\| \le D + T.$$

This gives the right-hand inclusion, which completes the proof of our claim.

We now take the measure of both sides and use the fact that the F_λ's are disjoint for distinct $\lambda \in \Lambda$ to get

$$\mu\bigl(B(T-D)\bigr) \le \sum_{\lambda \in \Lambda,\, \|\lambda\| \le T} \mu(F_\lambda) \le \mu\bigl(B(T+D)\bigr).$$

The measure μ is homogeneous and translation invariant, so in particular $\mu(F_\lambda) = \mu(F)$ and $\mu\bigl(B(T)\bigr) = \mu(B)T^r$ (where we write $B = B(1)$ for the unit ball). This gives

$$\mu(B)(T-D)^r \le \mu(F)\#\{\lambda \in \Lambda \mid \|\lambda\| \le T\} \le \mu(B)(T+D)^r.$$

Hence

$$\#\{\lambda \in \Lambda \mid \|\lambda\| \le T\} = \frac{\mu(B)}{\mu(F)} T^r + O(T^{r-1}).$$

Replacing $\|\lambda\|$ with $\sqrt{q(\lambda)}$ and T with $\sqrt{T}$ then gives the desired result, with the explicit value $a(q, \Lambda) = \mu(B)/\mu(F)$. □

The final result we will prove in this section is Mumford's estimate for the counting function of a curve C of genus $g \ge 2$. Assuming $C(k) \ne \emptyset$, we can embed $C(k)$ into its Jacobian $J(k)$. The Mordell–Weil theorem (C.0.1)

says that $J(k)$ is finitely generated, so Theorem B.6.3 tells us that the counting function of $J(k)$ looks like

$$N\big(J(k),T\big) = a(\log T)^{r/2} + O\big((\log T)^{(r-1)/2}\big).$$

Since $C(k) \subset J(k)$, we get an upper bound $N\big(C(k),T)\big) \ll (\log T)^{r/2}$. Prior to Faltings' proof of Mordell's conjecture, Mumford had shown that the points of $C(k)$ are widely dispersed in $J(k)$. More precisely, he proved that $N\big(C(k),T)\big) \ll \log\log T$. With the tools we have assembled, the proof of Mumford's estimate is not difficult, so we present it here. Despite the fact that Mumford's theorem has been superseded by Faltings' work, the proof is well worth studying, because Vojta and Bombieri use Mumford's ideas as the starting point in proving that $C(k)$ is actually finite.

Theorem B.6.5. (Mumford [1]) *Let C/k be a curve of genus $g \geq 2$ defined over a number field. Then there is a constant c, depending on C/k and the height function used for counting, such that*

$$N\big(C(k),T\big) \leq c\log\log T \qquad \text{for all } T \geq e^e.$$

PROOF. Fix a basepoint $P_0 \in C(\bar{k})$ and use it to embed C in its Jacobian in the standard way,

$$j : C \longrightarrow J, \qquad P \longmapsto \text{class}\big((P) - (P_0)\big).$$

Later we will take $P_0 \in C(k)$, but for the first part of the proof we do not need to make this assumption. We also let $\Theta = j(C) + \cdots + j(C)$ be the theta divisor on J, we let $\Delta \subset C \times C$ be the diagonal, and we define the usual maps

$$\begin{aligned} s_{12}, p_1, p_2 &: A \times A \longrightarrow A, \\ p_1, p_2 &: C \times C \longrightarrow C, \end{aligned} \qquad \begin{cases} s_{12}(x,y) = x + y, \\ p_1(x,y) = x, \\ p_2(x,y) = y. \end{cases}$$

The theta divisor Θ is ample (A.8.2.3(b)), and the same is true of the symmetric divisor $\Theta + [-1]^*\Theta$. The bilinear pairing

$$\langle\,\cdot\,,\,\cdot\,\rangle_\Theta : J(k) \times J(k) \longrightarrow \mathbb{R},$$
$$\langle x, y\rangle_\Theta = \frac{\hat{h}_{J,\Theta}(x+y) - \hat{h}_{J,\Theta}(x) - \hat{h}_{J,\Theta}(y)}{2},$$

depends only on the quadratic part of the canonical height (see B.5.6), so (B.5.3) tells us that it induces a positive definite pairing on $J(k) \otimes \mathbb{R}$.

Recall the fundamental divisor relation

$$(j \times j)^*(s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta) \sim -\Delta + p_1^*(P_0) + p_2^*(P_0)$$

from Theorem A.8.2.1. We use this relation and basic properties of the height machine to compute for any $P, Q \in C(\bar{k})$,

$$\begin{aligned}
2\langle j(P), j(Q)\rangle &= \hat{h}_{J,\Theta}\big(j(P)+j(Q)\big) - \hat{h}_{J,\Theta}\big(j(P)\big) - \hat{h}_{J,\Theta}\big(j(Q)\big) \\
&= h_{J,\Theta}\big(j(P)+j(Q)\big) - h_{J,\Theta}\big(j(P)\big) - h_{J,\Theta}\big(j(Q)\big) + O(1) \\
&= h_{J,\Theta}\big(s_{12}(j(P),j(Q))\big) \\
&\qquad - h_{J,\Theta}\big(p_1(j(P),j(Q))\big) - h_{J,\Theta}\big(p_2(j(P),j(Q))\big) + O(1) \\
&= h_{J\times J, s_{12}^*\Theta}\big(j(P),j(Q)\big) \\
&\qquad - h_{J\times J,p_1^*\Theta}\big(j(P),j(Q)\big) - h_{J\times J,p_2^*\Theta}\big(j(P),j(Q)\big) + O(1) \\
&= h_{C\times C,(j\times j)^*(s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta)}(P,Q) + O(1) \\
&= h_{C\times C,-\Delta+p_1^*(P_0)+p_2^*(P_0)}(P,Q) + O(1) \\
&= -h_{C\times C,\Delta}(P,Q) + h_{C,(P_0)}\big(p_1(P,Q)\big) \\
&\qquad + h_{C,(P_0)}\big(p_2(P,Q)\big) + O(1) \\
&= -h_{C\times C,\Delta}(P,Q) + h_{C,(P_0)}(P) + h_{C,(P_0)}(Q) + O(1).
\end{aligned}$$

The divisor Δ is effective, so (B.3.2e) tells us that $-h_{C\times C,\Delta}(P,Q)$ is bounded (above), provided that $P \neq Q$. We next need to relate $h_{C,(P_0)}$ back to the canonical height on J. To do this, we will use the divisor relation

$$j^*\Theta + j^*\Theta^- \sim 2g(P_0) + \kappa$$

provided by Theorem A.8.2.1, where $\Theta^- = [-1]^*\Theta$ and $\kappa \in \mathrm{Div}^0(C)$ is a certain divisor of degree 0. Using this formula and the height machine, we compute

$$\begin{aligned}
2gh_{C,(P_0)}(P) + h_{C,\kappa}(P) &= h_{C,2g(P_0)+\kappa}(P) + O(1) \\
&= h_{C,j^*\Theta+j^*\Theta^-}(P) + O(1) \\
&= h_{J,\Theta+\Theta^-}\big(j(P)\big) + O(1) \\
&= \hat{h}_{J,\Theta+\Theta^-}\big(j(P)\big) + O(1) \\
&= 2\langle j(P), j(P)\rangle_\Theta + O(1).
\end{aligned}$$

Now, the divisor (P_0) is an ample divisor on C, while κ is a divisor of degree 0. It follows from (B.3.5) that the ratio $h_{C,\kappa}(P)/h_{C,(P_0)}(P)$ goes to 0, so for any $\varepsilon > 0$ we have an estimate

$$\big|h_{C,\kappa}(P)\big| \le \varepsilon h_{C,(P_0)}(P) + O(1) \qquad \text{for all } P \in C(\bar{k}).$$

Of course, the $O(1)$ now also depends on ε. Substituting this in above, we find that

$$\begin{aligned}
2\langle j(P), j(P)\rangle_\Theta &= 2gh_{C,(P_0)}(P) + h_{C,\kappa}(P) + O(1) \\
&\ge (2g-\varepsilon)h_{C,(P_0)}(P) + O(1).
\end{aligned}$$

We can rewrite this as

$$h_{C,(P_0)}(P) \le \frac{1+\varepsilon}{g}\langle j(P), j(P)\rangle_\Theta + O(1) \qquad \text{for all } P \in C(\bar{k}),$$

where we have replaced ε by $\varepsilon/(2g-\varepsilon)$ to make the formula neater. Finally, using this in our formula for $\langle j(P), j(Q)\rangle$ given above, we obtain

$$\begin{aligned} 2\langle j(P), j(Q)\rangle_\Theta &= -h_{C\times C,\Delta}(P,Q) + h_{C,(P_0)}(P) + h_{C,(P_0)}(Q) + O(1) \\ &\le -h_{C\times C,\Delta}(P,Q) + \frac{1+\varepsilon}{g}\langle j(P), j(P)\rangle_\Theta \\ &\quad + \frac{1+\varepsilon}{g}\langle j(Q), j(Q)\rangle_\Theta + O(1). \end{aligned}$$

This estimate, which says that the points in $j(C)$ are "widely spaced" in J, is of sufficient importance to state as an independent proposition. Such an estimate is called a *gap principle*, because it says that there is a gap between solutions.

Proposition B.6.6. (Gap principle) *Let C/k be a curve of genus $g \ge 1$, let $P_0 \in C(\bar{k})$ be a fixed basepoint, let $\Delta \in \operatorname{Div}(C\times C)$ be the diagonal, and fix a height function $h_{C\times C,\Delta}$. Use P_0 to define an embedding $j : C \hookrightarrow J$ of C into its Jacobian, and let $\Theta \in \operatorname{Div}(J)$ be the theta divisor. Let*

$$\langle\cdot,\cdot\rangle_\Theta : J(\bar{k}) \times J(\bar{k}) \longrightarrow \mathbb{R}$$

be the canonical height pairing attached to Θ, and write

$$\|x\|_\Theta = \sqrt{\langle x, x\rangle_\Theta}$$

for the associated norm. Fix a constant $\varepsilon > 0$.

(a) *There is a constant c_1, depending on the above data, such that*

$$\langle j(P), j(Q)\rangle_\Theta \le \frac{1+\varepsilon}{2g}\left(\|j(P)\|_\Theta^2 + \|j(Q)\|_\Theta^2\right) - h_{C\times C,\Delta}(P,Q) + c_1 \qquad \text{for all } P,Q \in C(\bar{k}).$$

If we restrict to points with $P \ne Q$, then the $h_{C\times C,\Delta}(P,Q)$ term may be omitted.

(b) *There is a constant c_2, depending on the above data, such that*

$$\|j(P) - j(Q)\|_\Theta^2 \ge \left(1 - \frac{1+\varepsilon}{g}\right)\left(\|j(P)\|_\Theta^2 + \|j(Q)\|_\Theta^2\right) - c_2 \qquad \text{for all } P,Q \in C(\bar{k}) \text{ with } P \ne Q.$$

PROOF. (a) Up to a change of notation, we have already proven the first statement above. As for the second statement, we observe that Δ is an effective divisor, so Theorem B.3.2(e) tells us that $h_{C\times C,\Delta}$ is bounded below away from the base locus of the linear system $|\Delta|$. In particular, $-h_{C\times C,\Delta}$ is bounded above for all points not lying on Δ. (We remark that since $\Delta^2 = 2 - 2g < 0$, the linear system $|\Delta|$ consists only of Δ, so the base locus is exactly Δ. This explains why the $h_{C\times C,\Delta}(P,Q)$ term is necessary if $P = Q$.) This completes the proof of (a).

(b) To prove this part, we just use (a) and the identity

$$\langle x, y\rangle_\Theta = \frac{\|x\|_\Theta^2 + \|y\|_\Theta^2 - \|x-y\|_\Theta^2}{2}.$$

A little bit of algebra gives the desired result. □

Remark. (i) We amplify our earlier remarks on why Proposition B.6.6 is called a "gap principle." Abstractly, we have a real vector space V with a Euclidean norm $\|\cdot\|$ and a subset $S \subset V$. (In the situation of (B.6.6), $V = J(\bar{k}) \otimes \mathbb{R}$, $\|\cdot\| = \|\cdot\|_\Theta$, and S is the image of $j(C(\bar{k}))$ in V.) We are given positive constants α and β such that

$$\|x-y\|^2 \geq \alpha\bigl(\|x\|^2 + \|y\|^2\bigr) - \beta$$

for all $x, y \in S$ with $x \neq y$. Intuitively, this inequality means that if $\|x\|$ or $\|y\|$ is large, then x and y cannot be too close to each other. Or thinking about it another way, for each $x \in S$, there is a ball around x of radius $\sqrt{\alpha\|x\|^2 - \beta}$ that contains no other points of S. The larger the value of $\|x\|$, the larger the ball, so the points of S are forced further and further apart. When this intuition is quantified, it shows that the counting function

$$\bigl\{x \in S \bigm| \|x\| \leq T\bigr\}$$

grows very slowly as a function of T. The details are given in Lemma B.6.7 below.

(ii) We observe that everything we have done is valid for $g = 1$. In particular, the gap estimates in (B.6.6) hold for $g = 1$. However, when $g = 1$, these estimates are trivially true and give no additional information. This is especially true of (B.6.6(b)), which for $g = 1$ says that a nonnegative number is larger than a nonpositive number!

We now have the following general setup. We have a set $C(k)$ that we are trying to count, and we have an embedding of $C(k)$ into a group $J(k)$ on which we have a pairing $\langle\cdot,\cdot\rangle$. We further know that the points in $C(k)$ satisfy a gap condition as described in Proposition B.6.6. We now prove an abstract counting lemma that shows that sets satisfying such a gap condition are sparsely distributed.

Lemma B.6.7. *Let V be a finite-dimensional real vector space, let $\|\cdot\|$ be a Euclidean norm on V, let Λ be a lattice in V, and let $S \subset \Lambda$ be a subset of Λ. Suppose that there are constants $a, b > 0$ such that*

$$\|x-y\|^2 \geq a\bigl(\|x\|^2 + \|y\|^2\bigr) - b \qquad \text{for all } x, y \in S \text{ with } x \neq y.$$

Then there is a constant $c > 0$ such that

$$\#\bigl\{x \in S \bigm| \|x\| \leq T\bigr\} \leq c\log(T) \qquad \text{for all } T \geq 2.$$

Notice that Lemma B.6.4 says that

$$\#\bigl\{x \in \Lambda \bigm| \|x\| \leq T\bigr\} \sim T^{\dim(V)}.$$

Thus the gap condition on S leads to an exponential decrease in the number of points.

PROOF. For any numbers $u \geq v$, we let

$$S(u,v) = \bigl\{x \in S \bigm| u \leq \|x\| \leq v\bigr\}.$$

Further, for any $x \in V$ and any $r \geq 0$, we let $B_x(r)$ be the ball of radius r centered at x,

$$B_x(r) = \bigl\{z \in V \bigm| \|z-x\|^2 < r\bigr\}.$$

Suppose now that $u \geq \sqrt{b/a}$ and that $x, y \in S(u,v)$ are distinct points. Then the gap condition implies that

$$\|x-y\|^2 \geq a\bigl(\|x\|^2 + \|y\|^2\bigr) - b \geq 2au^2 - b \geq au^2.$$

In other words, the distance from x to y is at least $u\sqrt{a}$, and hence

$$B_x\left(\frac{1}{2}u\sqrt{a}\right) \cap B_y\left(\frac{1}{2}u\sqrt{a}\right) = \emptyset.$$

On the other hand, we clearly have

$$B_x\left(\frac{1}{2}u\sqrt{a}\right) \subset B_0\left(\|x\| + \frac{1}{2}u\sqrt{a}\right) \subset B_0\left(v + \frac{1}{2}u\sqrt{a}\right),$$

since $\|x\| \leq v$ by our assumption that $x \in S(u,v)$. It follows that the large ball $B_0\left(v + \frac{1}{2}u\sqrt{a}\right)$ contains the disjoint union of the balls $B_x\left(\frac{1}{2}u\sqrt{a}\right)$ as x ranges over $S(u,v)$, so we obtain a volume inequality

$$\begin{aligned}\operatorname{Vol} B_0\left(v + \frac{1}{2}u\sqrt{a}\right) &\geq \operatorname{Vol}\Bigl(\bigcup_{x \in S(u,v)} B_x\left(\frac{1}{2}u\sqrt{a}\right)\Bigr) \\ &\geq \sum_{x \in S(u,v)} \operatorname{Vol}\left(B_x\left(\frac{1}{2}u\sqrt{a}\right)\right).\end{aligned}$$

But for any $z \in V$, the volume of a ball of radius r is

$$\mathrm{Vol}\big(B_z(r)\big) = r^n \,\mathrm{Vol}\big(B_0(1)\big),$$

where $n = \dim V$. Substituting this and canceling the volume of the unit ball gives

$$\left(v + \frac{1}{2}u\sqrt{a}\right)^n \geq \#S(u,v) \cdot \left(\frac{1}{2}u\sqrt{a}\right),$$

and hence

$$\#S(u,v) \leq \left(\frac{2v}{u\sqrt{a}} + 1\right)^n.$$

Remember that we have proved this estimate only under the assumption that $u \geq \sqrt{b/a}$. To ease notation, we will let $\alpha = \sqrt{b/a}$, and we will apply the estimate for $S(u,v)$ with

$$u = \alpha e^i \qquad \text{and} \qquad v = \alpha e^{i+1} \qquad \text{as } i \text{ ranges over } 0 \leq i \leq \log(T/\alpha).$$

This gives the bound

$$\begin{aligned}\#S(\alpha,T) &= \sum_{0 \leq i \leq \log(T/\alpha)} \#S(\alpha e^i, \alpha e^{i+1}) \\ &\leq \sum_{0 \leq i \leq \log(T/\alpha)} \left(\frac{2e}{\sqrt{a}} + 1\right)^n \leq \left(1 + \log\frac{T}{\alpha}\right)\left(\frac{2e}{\sqrt{a}} + 1\right)^n.\end{aligned}$$

Further, the lattice Λ has only finitely many elements of bounded norm, so $\#S(0,\alpha)$ is clearly finite. Hence

$$\#S(0,T) \leq \#S(0,\alpha) + \#S(\alpha,T) \leq c_1 \log(T) + c_2,$$

where the constants c_1, c_2 are independent of T. We can even omit the c_2 if we assume that $T \geq 2$ and take a larger c_1. This completes the proof of Lemma B.6.7. □

Mumford's Theorem B.6.5 is now an immediate consequence of the gap principle (B.6.6) and the counting lemma (B.6.7).

PROOF (Mumford's theorem (B.6.5)). Our first observation is that we can use any convenient height function in order to compute the counting function $N\big(C(k),T\big)$. To see this, let H_D and H_E be Weil heights on C with respect to ample divisors D and E, respectively. The assumption that D and E are ample implies that they have positive degrees (A.4.2.4), and then (B.3.5) tells us that

$$\lim_{\substack{P \in C(\bar{k}) \\ h_D(P) \to \infty}} \frac{\log H_D(P)}{\log H_E(P)} = \frac{\deg D}{\deg E}.$$

It follows that there are constants $c_1, c_2 > 0$ such that

$$N\big(C(k), H_E, T\big) \le c_1 N\big(C(k), H_D, T\big) - c_2.$$

Hence if (B.6.5) is true for the counting function $N\big(C(k), H_D, T\big)$, then it is also true for the counting function $N\big(C(k), H_E, T\big)$.

If $C(k)$ is empty, there is nothing to prove. So we can assume that there is at least one point $P_0 \in C(k)$, and we use P_0 to embed C in its Jacobian, $j : C \hookrightarrow J$, as usual. We let $\langle\cdot,\cdot\rangle_\Theta$ and $\|\cdot\|_\Theta$ be the canonical height pairing and associated norm attached to the divisor $\Theta \in \operatorname{Div}(J)$. We will compute $N\big(C(k), T\big)$ using a height function associated to the divisor $\frac{1}{2}j^*(\Theta + \Theta^-)$. More precisely, we will count

$$N\big(C(k), T\big) = \#\big\{P \in C(k) \bigm| \|j(P)\|_\Theta^2 \le \log T\big\}.$$

(Notice that we have to put in $\log T$, since $N\big(C(k), T\big)$ is computed using a multiplicative height, while $\|\cdot\|_\Theta^2$ is a logarithmic height.)

Let $V = J(k) \otimes \mathbb{R}$, let Λ be the image of $J(k)$ in V, and let S be the image of $C(k)$ in V. We know from the Mordell–Weil theorem (C.0.1) that V is a finite-dimensional vector space, and (B.5.3) tells us that the norm $\|\cdot\|_\Theta$ on $J(k)$ induces a Euclidean norm $\|\cdot\|$ on V. Further, the gap principle (B.6.6b) says that

$$\|x-y\|^2 \ge \left(1 - \frac{1+\varepsilon}{g}\right)\big(\|x\|^2 + \|y\|^2\big) - c_3 \qquad \text{for all } x, y \in S \text{ with } x \ne y.$$

We are assuming that the genus g of C is at least 2, so taking $\varepsilon = \frac{1}{2}$, we find that

$$\|x-y\|^2 \ge \tfrac{1}{4}\big(\|x\|^2 + \|y\|^2\big) - c_3 \qquad \text{for all } x, y \in S \text{ with } x \ne y.$$

We are in exactly the situation to apply the counting lemma (B.6.7), which gives the estimate

$$\#\big\{x \in S \bigm| \|x\| \le \log T\big\} \le c_4 \log\log T \qquad \text{for all } T \ge 8.$$

Finally, we just need to observe that the kernel of the map $J(k) \to J(k) \otimes \mathbb{R} = V$ is exactly the torsion subgroup of $J(k)$. In particular, this map is finite-to-one, so the same is true of the map $C(k) \to S$. Precisely, we have

$$N\big(C(k), T\big) \le \#J(k)_{\text{tors}} \cdot \#\big\{x \in S \bigm| \|x\| \le \log T\big\}.$$

This estimate, combined with the earlier inequality, completes the proof of Mumford's theorem (B.6.5). □

It is quite instructive to compare the orders of growth of the counting function $N\big(V(k),T\big)$ given by Theorems B.6.2, B.6.3, and B.6.5. At the coarsest level, they say that $N\big(V(k),T\big)$ grows polynomially for projective space, logarithmically for abelian varieties, and at most like $\log\log$ for curves of genus at least 2. A slightly weaker way of stating these facts is given by the following limits:

$$\lim_{T\to\infty}\frac{\log\log N\big(\mathbb{P}^n(k),T\big)}{\log\log T}=1, \qquad \text{(projective space)}$$
$$\lim_{T\to\infty}\frac{\log\log N\big(A(k),T\big)}{\log\log\log T}=1, \qquad \text{(abelian variety with } \#A(k)=\infty)$$
$$\limsup_{T\to\infty}\frac{\log\log N\big(C(k),T\big)}{\log\log\log\log T}\le 1. \qquad \text{(curve of genus } g(C)\ge 2)$$

The reason we have taken $\log\log N\big(V(k),T\big)$ is that its order of growth is independent of the choice of height function used for counting. (See Exercise B.15.) Further, the use of $\log\log$ has a smoothing effect. This suggests that we consider the possible growth orders for the $\log\log$ counting function

$$\log\log N\big(V(k),T\big).$$

Projective spaces and abelian varieties give examples for which it grows like $\log\log T$ and $\log\log\log T$, respectively. Notice that Mumford's theorem gives only an upper bound for $\log\log N\big(C(k),T\big)$, and in fact this upper bound is not sharp, since Faltings' theorem says that $N\big(C(k),T\big)$ is actually bounded independently of T. All of this leads to the following questions.

Question B.6.8. *What are the possible behaviors for the counting function* $\log\log N\big(V(k),T\big)$*? For example, let* V/k *be a projective variety, and let* $U\subset V$ *be a Zariski open subset. Is it true that the counting function for* $U(k)$ *must satisfy one of the following conditions?*

$$\lim_{T\to\infty}\frac{\log\log N\big(U(k),T\big)}{\log\log T}=1, \qquad \textit{(polynomial growth)}$$
$$\lim_{T\to\infty}\frac{\log\log N\big(U(k),T\big)}{\log\log\log T}=1, \qquad \textit{(logarithmic growth)}$$
$$\log\log N\big(U(k),T\big) \textit{ is bounded as } T\to\infty. \qquad \textit{(bounded growth)}$$

Notice that if question (B.6.8) has an affirmative answer, then Mumford's theorem (B.6.5) would imply the finiteness of $C(k)$. It is possible to formulate similar questions for S-integer points on affine varieties. See Silverman [4] for further details.

The behavior of $\log\log N\big(V(k),T\big)$ is an extremely coarse measure of the distribution of rational points on $V(k)$. At the other extreme, Batyrev

and Manin have formulated very precise conjectures for the order of growth of $N\big(V(k),T\big)$ in certain cases. We give two examples here. See Section F.5 for a further discussion, and Batyrev and Manin [1] and Franke, Manin, and Tschinkel [1] for additional information.

Conjecture B.6.9. (Batyrev–Manin [1]) *Let V/k be a smooth projective variety, and let K_V be a canonical divisor on V.*
(a) *Suppose that the anticanonical divisor $-K_V$ is ample. Then there is an integer $t \ge 0$ and a Zariski open subset $U \subset V$ such that possibly after replacing the field k by a finite extension, we have*

$$N\big(U(k), H_{-K_V}, T\big) \sim cT(\log T)^t \qquad \text{as } T \to \infty.$$

(Here H_{-K_V} is the height relative to the field k, not the absolute height.)
(b) *Suppose that some nonzero multiple of K_V is linearly equivalent to 0. Then for every $\varepsilon > 0$ there is a nonempty Zariski open subset $U_\varepsilon \subset V$ such that*

$$N\big(U_\varepsilon(k), T\big) \le T^\varepsilon \qquad \text{for all sufficiently large } T.$$

B.7. Heights and Polynomials

In this section we prove some elementary height estimates for polynomials that will be used in Parts D and E. The reader may wish to skip this section until it is needed.

The (*affine*) *height* of a polynomial is defined to be the height of its coefficients taken as affine coordinates. Thus writing

$$f = \sum_{i=(i_1,\dots,i_n)\in I} a_i x_1^{i_1} \cdots x_n^{i_n},$$

the (absolute affine) height of f is

$$h(f) = h\big([1, \dots, a_i, \dots]_{i\in I}\big).$$

Alternatively, if we define the *Gauss norm* of a polynomial f with respect to an absolute value v to be

$$|f|_v = \max_{i\in I} |a_i|_v,$$

then

$$H_k(f) = \prod_{v\in M_k} \max\{1, |f|_v^{n_v}\}$$

and

$$h(f) = \log H(f) = \frac{1}{[k:\mathbb{Q}]} \sum_{v \in M_k} n_v \log \max\{1, |f|_v\}.$$

For example,

$$h(6x^2 + 3xy + 12y) = h([1,6,3,12]) = \log(12).$$

More generally, if $\mathcal{F} = \{f_1, \ldots, f_r\}$ is a collection of polynomials, then we define the height of the collection to be

$$h(\mathcal{F}) = \max\{h(f_1), \ldots, h(f_r)\}.$$

Remark B.7.0. For some applications, it is more convenient to use the (*projective*) *height* of a polynomial, which we define to be the height of its coefficients taken as homogeneous coordinates. Thus

$$H_k(f) = \prod_{v \in M_k} |f|_v^{n_v} \quad \text{and} \quad h(f) = \frac{1}{[k:\mathbb{Q}]} \sum_{v \in M_k} n_v \log |f|_v.$$

The projective height of the above example is

$$h(6x^2 + 3xy + 12y) = h([6,3,12]) = h([2,1,4]) = \log(4).$$

We will mostly be using affine polynomial heights, and we will specify when this is not the case.

Proposition B.7.1. *Let k be a number field and $F \in k[x_0, \ldots, x_n]$ a homogeneous polynomial of degree d, say*

$$F(x) = F(x_0, \ldots, x_n) = \sum_{\substack{i=(i_0,\ldots,i_n) \\ i_0+\cdots+i_n=d}} a_i x_0^{i_0} x_1^{i_1} \cdots x_n^{i_n},$$

and let $x = (x_0, \ldots, x_n) \in \bar{k}^{n+1}$.
(a) *Let v be an absolute value on k, extended in some way to $\bar{k}$, and let $\nu_v(F) = \binom{n+d}{n}$ if v is archimedean and $\nu_v(F) = 1$ otherwise. Then*

$$\left|F(x)\right|_v \le \nu_v(F)\bigl(\max_i |a_i|_v\bigr)\bigl(\max_j |x_j|_v\bigr)^d.$$

(b) $$h\bigl(F(x)\bigr) \le dh(x) + h(F) + \min\{n\log(n+d), (n+d)\log 2\}.$$

Proof. (a) The desired estimate is immediate from the triangle inequality applied to the sum $F(x)$, once we observe that there are $\binom{n+d}{n}$ terms in the sum.
(b) Taking the logarithm of (a) and summing over all places of k gives (b). We have also used the trivial estimate

$$\binom{n+d}{n} \le \min\left\{(n+d)^n, 2^{n+d}\right\}.$$

□

Next we give some elementary estimates for the heights of sums and products of polynomials.

Proposition B.7.2. *Let $\mathcal{F} = \{f_1, \dots, f_r\}$ be a collection of polynomials in $k[X_1, \dots, X_m]$, where k is a number field.*
(a) *Let $\deg f_i$ be the total degree of f_i. Then*

$$h(f_1 f_2 \cdots f_r) \le \sum_{i=1}^{r} \bigl(h(f_i) + (\deg f_i + m)\log 2\bigr)$$
$$\le r \max_{1\le i\le r} \bigl\{h(f_i) + (\deg f_i + m)\log 2\bigr\}.$$

(This estimate will be slightly refined in Proposition B.7.4 below.)
(b)

$$h(f_1 + f_2 + \cdots + f_r) \le \sum_{i=1}^{r} h(f_i) + \log r.$$

(c) *Suppose that $f_1, \dots, f_r \in R[X_1, \dots, X_m]$ have coefficients in the ring of integers R of k. Then*

$$h(f_1 + f_2 + \cdots + f_r) \le [k : \mathbb{Q}]h(\mathcal{F}) + \log r.$$

(This estimate is useful when k is fixed and r is large.)

PROOF. Let $f_i = \sum_E a_{iE} X^E$, where $E = (E_1, \dots, E_m)$ runs over m-tuples of nonnegative integers and $X^E = X_1^{E_1} \cdots X_m^{E_m}$. We then have

$$f_1 \cdots f_r = \sum_E \Bigl(\sum_{e_1+\cdots+e_r=E} a_{1e_1} \cdots a_{re_r} \Bigr) X^E,$$

and hence for any $v \in M_k$,

$$|f_1 \cdots f_r|_v = \max_E \Bigl| \sum_{e_1+\cdots+e_r=E} a_{1e_1} \cdots a_{re_r} \Bigr|_v.$$

Let N be an upper bound for the number of nonzero terms in the sums, and as usual let $N_v = 1$ for v nonarchimedean and $N_v = N$ for v archimedean. We then have

$$|f_1 \cdots f_r|_v \le \max_E \Bigl(N_v \max_{e_1+\cdots+e_r=E} |a_{1e_1} \cdots a_{re_r}|_v \Bigr)$$
$$\le N_v \prod_{i=1}^{r} \max_{e_i} \{1, |a_{ie_i}|_v\}$$
$$\le N_v \prod_{i=1}^{r} \max\{1, |f_i|_v\}.$$

Raising to the $n_v/[k:\mathbb{Q}]$ power and taking the product over all v then gives

$$\begin{aligned} H(f_1 \cdots f_r) &= \prod_{v \in M_k} \max\left\{1, |f_1 \cdots f_r|_v^{n_v/[k:\mathbb{Q}]}\right\} \\ &\le \prod_{v \in M_k} \left\{N_v \prod_{i=1}^{r} \max\{1, |f_i|_v\}\right\}^{n_v/[k:\mathbb{Q}]} \\ &\le N \prod_{i=1}^{r} H(f_i). \end{aligned}$$

(Recall that $\sum_{v \in M_k^\infty} n_v = [k:\mathbb{Q}]$.) To find an admissible value for N, we proceed as follows. We note that the number of r-tuples of nonnegative integers with sum equal to E_i is $\binom{E_i+r-1}{E_i}$, so the number of terms we are trying to estimate is smaller than

$$\max_E \prod_{i=1}^{m} \binom{E_i + r - 1}{E_i} \le \max_E \prod_{i=1}^{m} \left(2^{E_i+r-1}\right) \le 2^{\deg(f_1 \cdots f_r) + m(r-1)}.$$

Hence we may choose

$$\log N = \log 2 \left(\sum_{i=1}^{r} (\deg f_i + m)\right),$$

which completes the proof of (a).

(b) Keeping the same notation for the coefficients of the f_i's, we have

$$f_1 + \cdots + f_r = \sum_e (a_{1e} + \cdots + a_{re}) X^e.$$

Thus for any $v \in M_k$,

$$|f_1 + \cdots + f_r|_v = \max_e |a_{1e} + \cdots + a_{re}|_v.$$

Writing $r_v = 1$ or $r_v = r$ for v nonarchimedean or archimedean as usual, we get

$$\max\{1, |f_1 + \cdots + f_r|_v\} \le r_v \max_{i,e}\{1, |a_{ie}|\} \le r_v \prod_{i=1}^{r} \max_e\{1, |a_{ie}|\}.$$

Raising to the $n_v/[k:\mathbb{Q}]$ power and taking the product over v then gives

$$H(f_1 + \cdots + f_r) \le r \prod_{i=1}^{r} H(f_i).$$

(c) Since we are assuming that the f_i's have algebraic integer coefficients, the same will be true of $f_1 + \cdots + f_r$. This implies that for any nonarchimedean v, we have

$$\max\{1, |f_1 + \cdots + f_r|_v\} = \max\{1, |f_1|_v\} = \cdots = \max\{1, |f_r|_v\} = 1.$$

This implies that only the archimedean places contribute to the height of $f_1 + \cdots + f_r$, so

$$\begin{aligned} H_k(f_1 + \cdots + f_r) &= \prod_{v \in M_k^\infty} \max\{1, |f_1 + \cdots + f_r|_v^{n_v}\} \\ &\le \prod_{v \in M_k^\infty} \left(r \cdot \max_{1 \le i \le r} \{1, |f_i|_v^{n_v}\} \right) \\ &\le r^{[k:\mathbb{Q}]} \max_{1 \le i \le r} \max_{v \in M_k^\infty} \{1, |f_i|_v^{n_v}\}^{[k:\mathbb{Q}]} \\ &\le r^{[k:\mathbb{Q}]} \max_{1 \le i \le r} H_k(f_i)^{[k:\mathbb{Q}]}. \end{aligned}$$

(We are using here the fact that $\#M_k^\infty \le [k : \mathbb{Q}]$.) Taking $[k : \mathbb{Q}]$th roots gives the desired result:

$$H(f_1 + \cdots + f_r) \le r \max_{i=1}^{r} H(f_i)^{[k:\mathbb{Q}]}. \qquad \square$$

The next inequality will be used (among other places) in the proof of Roth's theorem.

Proposition B.7.3. (Gelfand's inequality). *This proposition uses projective polynomial heights. Let $d_1, \ldots, d_r$ be integers, and let $f_1, \ldots, f_r \in \bar{\mathbb{Q}}[X_1, \ldots, X_m]$ be polynomials whose product satisfies $\deg_{X_i}(f_1 \cdots f_r) \le d_i$ for each $1 \le i \le r$. Then*

$$\sum_{i=1}^{r} h(f_i) \le h(f_1 \cdots f_r) + d_1 + \cdots + d_m.$$

Remark 7.3.1. As a first approach, if we use projective space to parametrize the set of polynomials of given degrees in m variables, then the map that sends an r-tuple of polynomials $(f_1, \ldots, f_r)$ to their product $f_1 \cdots f_r$ becomes a rational map

$$\phi : \mathbb{P}^{M_1} \times \cdots \times \mathbb{P}^{M_r} \to \mathbb{P}^N.$$

This map may be described as the composition of a Segre embedding with a linear projection whose center is disjoint from the image of the Segre

embedding. (The fact that the center and the image are disjoint is because the product of nonzero polynomials is again nonzero.) Then the height machine (B.3.2) immediately gives the estimate

$$h(f_1\cdots f_r) = h(f_1) + \cdots + h(f_r) + O(1).$$

So the content of Gelfand's inequality is to give an explicit bound for the $O(1)$ term. We also note that a converse inequality of the form $H(f) \le c(d_1,\ldots,d_m)\prod_{i=1}^r H(f_i)$ can be easily established using only the triangle inequality; cf. Proposition B.7.2.

PROOF. We start by recalling Gauss's lemma, which states in this context that

$$|f_1\cdots f_r|_v = |f_1|_v\cdots|f_r|_v \quad \text{for all nonarchimedean } v.$$

The crux of the proof of Gelfand's inequality is the proof of an analogous archimedean estimate,

$$\prod_{i=1}^r |f_i| \le e^{d_1+\cdots+d_m}|f|, \tag{$*$}$$

valid for all polynomials $f, f_1,\ldots,f_r \in \mathbb{C}[X_1,\ldots,X_m]$. Indeed, granting $(*)$ for a moment, we can compute

$$\begin{aligned}\prod_{i=1}^r H_k(f_i) &= \prod_{i=1}^r\prod_{v\in M_k}|f_i|_v^{n_v}\\ &\le \prod_{v\in M_k^0}|f_1\cdots f_r|_v^{n_v}\prod_{v\in M_k^\infty}e^{n_v(d_1+\cdots+d_m)}|f_1\cdots f_r|_v^{n_v}\\ &\le e^{[k:\mathbb{Q}](d_1+\cdots+d_m)}H_k(f_1\cdots f_r),\end{aligned}$$

and then taking $[k:\mathbb{Q}]$th roots gives Gelfand's inequality.

We will prove $(*)$ by introducing a multiplicative norm and an L^2-norm on the space of polynomials and comparing them to the Gauss norm. These norms, especially the Mahler measure, are interesting in their own rights and have been much studied.

Definition. Let

$$I=[0,1],\quad i=(i_1,\ldots,i_m),\quad t=(t_1,\ldots,t_m),\quad dt = dt_1\cdots dt_m,$$

and let $\mathbf{e}(t) = \big(\exp(2\pi i t_1),\ldots,\exp(2\pi i t_m)\big)$. For any complex polynomial

$$f=\sum_i a_i X_1^{i_1}\cdots X_m^{i_m} \in \mathbb{C}[X_1,\ldots,X_m],$$

we define the *Mahler measure of f* to be the quantity

$$M(f) = \exp\left(\int_{I^m} \log\bigl|f(\mathbf{e}(t))\bigr| dt\right),$$

and the *L^2-norm of f* to be the quantity

$$L_2(f) = \left(\int_I |f(\mathbf{e}(t))|^2 dt\right)^{1/2} = \left(\sum_i |a_i|^2\right)^{1/2}.$$

The inequality we are attempting to prove will follow from a series of comparisons between various norms. We start with the "easy" estimates.

Lemma B.7.3.1. *Let $f, g \in \mathbb{C}[X_1, \ldots, X_m]$ be polynomials, and suppose that $\deg_{X_j}(f) \le d_j$. Then*

(i) $L_2(f) \le [(d_1+1)\cdots(d_n+1)]^{1/2}|f|$.
(ii) $M(fg) = M(f)M(g)$.
(iii) $M(f) \le L_2(f)$.

PROOF. The first two formulas are straightforward. Indeed the number of coefficients of f is bounded by $(d_1+1)\cdots(d_m+1)$, and hence

$$\sum_i |a_i|^2 \le (d_1+1)\cdots(d_m+1)\max|a_i|^2$$

Now (i) follows by taking square roots. Next, formula (ii) is immediate using the linearity of integration and the relationship between log and exp.

The third inequality deserves more explanation. We use Jensen's inequality, which states that if U is a set of measure 1 and if ϕ is a convex function, then

$$\phi\left(\int_U \psi\, d\mu\right) \le \int_U (\phi \circ \psi) d\mu.$$

Applying this with $U = I^m$, $d\mu = dt$, $\phi = \exp$, and $\psi(t) = 2\log|f(\mathbf{e}(t))|$, we get

$$M(f)^2 \le \int_{I^m} |f(\mathbf{e}(t))|^2 dt = L_2(f)^2,$$

and hence $M(f) \le L_2(f)$. □

Next we observe that for a polynomial in one variable

$$f = a_0 + a_1X + \cdots + a_dX^d = a_d(X-\alpha_1)\cdots(X-\alpha_d),$$

we have the formula

$$M(f) = |a_d|\prod_{i=1}^{d}\max\{1, |\alpha_i|\}.$$

In particular, if α is an algebraic number and if we take f to be the minimal polynomial of α, then this formula says that the height $H_k(\alpha)$ of α is equal to the Mahler measure $M(f)$ of its minimal polynomial. We also observe that using the multiplicativity of $M(f)$, this equality is equivalent to the well-known formula (Exercise B.19)

$$\int_0^1 \log|\alpha - e^{2i\pi t}|dt = \log\max\{1, |\alpha|\}.$$

We now use the equality $H_k(\alpha) = M(f)$ to prove a key estimate for the coefficients of a polynomial in terms of its Mahler measure.

Lemma B.7.3.2. *For any polynomial* $f = \sum_{i=0}^{d} a_i X^i \in \mathbb{C}[X]$,

$$|a_j| \le \binom{d}{j} M(f) \le 2^d M(f).$$

More generally, let

$$f = \sum_{\substack{0\le j_1\le d_1\\ \vdots \\ 0\le j_m\le d_m}} a_{j_1,\ldots,j_m} X_1^{j_1}\cdots X_m^{j_m} \in \mathbb{C}[X_1,\ldots,X_m]$$

be a polynomial satisfying $\deg_{X_h} f \le d_h$. *Then*

$$|a_{j_1,\ldots,j_m}| \le \binom{d_1}{j_1}\cdots\binom{d_m}{j_m} M(f) \le 2^{d_1+\cdots d_m} M(f).$$

PROOF. The proof is by induction on the number m of variables. For $m = 1$ we factor $f(X) = a_d \prod (X - \alpha_i)$. Then

$$|a_j| = |a_d| \Bigl| \sum_{h_1<\cdots<h_{d-j}} \alpha_{h_1}\cdots\alpha_{h_{d-j}} \Bigr| \le \binom{d}{j} |a_d| \prod_{i=1}^{d} \max\{1, |\alpha_i|\} = \binom{d}{j} M(f).$$

To assist in the induction, we set some notation. For any $1 \le n \le m$, we let

$$f_{k_1,\ldots,k_n}(X_{n+1},\ldots,X_m) = \sum_{h_{n+1}=0}^{d_{n+1}} \cdots \sum_{h_m=0}^{d_m} a_{k_1,\ldots,k_n,h_{m+1},\ldots,h_m} X_{n+1}^{h_{n+1}} \cdots X_m^{h_m}$$

with the convention that $f_{k_1,\ldots,k_m} = a_{k_1,\ldots,k_m}$ in the case $n = m$. This allows us to write

$$f(X_1,\ldots,X_m) = \sum_{k_1=0}^{d_1} f_{k_1}(X_2,\ldots,X_m) X_1^{k_1}, \qquad (*)_1$$

and more generally

$$f_{k_1,\ldots,k_{n-1}}(X_n,\ldots,X_m)=\sum_{k_n=0}^{d_n} f_{k_1,\ldots,k_n}(X_{n+1},\ldots,X_m)X_n^{k_n} \qquad (*)_n$$

From $(*)_1$ and the previous lemma in one variable, we deduce that for all $x_2,\ldots,x_m\in\mathbb{C}$,

$$\left|f_{k_1}(x_2,\ldots,x_m)\right|\le\binom{d_1}{k_1}\exp\left(\int_0^1\log\left|f(e^{2\pi it},x_2,\ldots,x_m)\right|dt\right).$$

Now we take the logarithm of both sides, evaluate at $(x_2,\ldots,x_m)=(e^{2\pi it_2},\ldots,e^{2\pi it_m})$, and integrate over $0\le t_2,\ldots,t_m\le 1$ to obtain

$$\begin{aligned}\log M(f_{k_1})&=\int_{I^{m-1}}\log\left|f_{k_1}(e^{2\pi it_2},\ldots,e^{2\pi it_m})\right|dt_2\cdots dt_m\\&\le\log\binom{d_1}{k_1}+\int_{I^m}\log\left|f(e^{2\pi it_1},\ldots,e^{2\pi it_m})\right|dt_1\cdots dt_m\\&\le\log\binom{d_1}{k_1}+\log M(f).\end{aligned}$$

This gives the inequality $M(f_{k_1})\le\binom{d_1}{k_1}M(f)$, and more generally, starting from formula $(*)_n$ and using the same argument, we get

$$M(f_{k_1,\ldots,k_n})\le\binom{d_n}{k_n}M(f_{k_1,\ldots,k_{n-1}}).$$

This gives the bound $|a_{k_1,\ldots,k_m}|\le\binom{d_m}{k_m}M(f_{k_1,\ldots,k_{m-1}})$ for the coefficients, and now the claim follows by putting together these inequalities. □

Let $\mu(f)$ denote the number of variables $X_1,\ldots,X_m$ that genuinely appear in f. Then using the trivial estimate

$$\binom{d}{k}\le 2^{d-1},\quad\text{valid for } d\ge 1,$$

we obtain

$$|f|\le 2^{d_1+\cdots+d_m-\mu(f)}M(f).$$

We can now finish the proof of Gelfand's inequality (B.7.3). We let $d_{ij}=\deg_{X_j}f_i$ and $d_j=\deg_{X_j}f$, so

$$d_j=\sum_{i=1}^r d_{ij}\qquad\text{and}\qquad\mu(f)\le\sum_{i=1}^r\mu(f_i).$$

Then

$$\begin{aligned}|f_1|\cdots|f_r| &\le \prod_{i=1}^{r}\left(2^{d_{i1}+\cdots+d_{im}-\mu(f_i)}M(f_i)\right)\\ &= 2^{d_1+\cdots+d_m-\sum_i \mu(f_i)}M(f)\\ &\le 2^{d_1+\cdots+d_m-\mu(f)}\left((d_1+1)\cdots(d_m+1)\right)^{1/2}|f|.\end{aligned}$$

Now observe that

$$2^d\sqrt{d+1} \le e^d \quad \text{for } d\ge 2 \text{ and for } d=0,$$

while if $d_i = 1$, then the X_i variable contributes to $\mu(f)$. This lets us simplify the inequality to obtain

$$|f_1|\cdots|f_r| \le e^{d_1+\cdots+d_m}|f|,$$

which completes the proof of Proposition B.7.3. □

The next result will be used in Part E, specifically in the proof of Eisenstein's estimate in Section E.9.

Proposition B.7.4. *Let k be a number field and v an absolute value on k. We write $\deg f$ for the total degree of a polynomial f, and for any integer N and absolute value v we set as usual*

$$N_v = \begin{cases} |N|_v & \text{if } v \text{ is archimedean,}\\ 1 & \text{if } v \text{ is nonarchimedean.}\end{cases}$$

Let $f, f_1, \ldots, f_r \in k[X_1,\ldots,X_n]$ be polynomials.
(a)

$$\left|\prod_{i=1}^{r} f_i\right|_v \le \min\left\{\prod_{i=2}^{r}(2\deg f_i)_v^n,\ \prod_{i=2}^{r}2_v^{\deg f_i}\right\}\times\prod_{i=1}^{r}|f_i|_v.$$

Note that the bound for the ratio $\left|\prod f_i\right|_v\Big/\prod|f_i|_v$ does not depend on f_1, but only on $f_2,\ldots,f_r$. This is often useful for induction arguments, where one of the f_i's may have much larger degree than the others.

(b)
$$\left|\sum_{i=1}^{r} f_i\right|_v \le r_v \max_{1\le i\le r}|f_i|_v.$$

(c)
$$\left|\frac{\partial f}{\partial X_j}\right|_v \le (\deg f)_v|f|_v.$$

(d) *Let* $b = (b_1, \ldots, b_n) \in k^n$, *and write* $|b|_v = \max |b_i|_v$. *Then*

$$\left|f(b)\right|_v \leq \min\left\{(2\deg f)_v^n, 2_v^{\deg f}\right\} \cdot |f|_v \cdot \max\left\{1, |b|_v\right\}^{\deg f}.$$

(e) *Let* b *and* $|b|_v$ *be as in (d), and define a shifted polynomial* $f_b(X) = f(X+b) = f(X_1+b_1, \ldots, X_n+b_n)$. *Then*

$$|f_b|_v \leq 2_v^{2\deg f} |f|_v \max\left\{1, |b|_v\right\}^{\deg f}.$$

PROOF. For a polynomial $f \in k[X_1, \ldots, X_n]$, we write

$$f(X) = \sum_e a_e X^e = \sum_{e_1=0}^{d_1} \cdots \sum_{e_n=0}^{d_n} a_e X_1^{e_1} \cdots X_n^{e_n},$$

where e is the multi-index $e = (e_1, \ldots, e_n)$ and $d_j = \deg_{X_j} f$. We observe that the number of nonzero monomials appearing in f satisfies

$$\begin{aligned}
(\#\text{ of nonzero } a_e\text{'s}) &\leq \prod_{j=1}^n (d_j+1) \\
&\leq \min\left\{\prod_{j=1}^n 2^{d_j},\ \prod_{j=1}^n (2\deg f)\right\} \\
&= \min\left\{2^{\deg f},\ (2\deg f)^n\right\}.
\end{aligned} \tag{1}$$

Now write $f_1, \ldots, f_r$ as $f_i = \sum_e a_{ie} X^e$. Then

$$\begin{aligned}
\prod_{i=1}^r f_i &= \prod_{i=1}^r \left(\sum_e a_{ie} X^e\right) \\
&= \sum_{e^{(1)}, \ldots, e^{(r)}} a_{1e^{(1)}} \cdots a_{re^{(r)}} X^{e^{(1)}+\cdots+e^{(r)}} \\
&= \sum_E \Bigl(\sum_{e^{(1)}+\cdots+e^{(r)}=E} a_{1e^{(1)}} \cdots a_{re^{(r)}}\Bigr) X^E.
\end{aligned} \tag{2}$$

We fix a multi-index $E = (E_1, \ldots, E_n)$ and look at the coefficient of X^E in (2):

$$\begin{aligned}
&\sum_{e^{(1)}+\cdots+e^{(r)}=E} a_{1e^{(1)}} \cdots a_{re^{(r)}} \\
&\qquad = \sum_{e_1^{(1)}+\cdots+e_1^{(r)}=E_1} \ \sum_{e_2^{(1)}+\cdots+e_2^{(r)}=E_2} \cdots \sum_{e_n^{(1)}+\cdots+e_n^{(r)}=E_n} a_{1e^{(1)}} \cdots a_{re^{(r)}}.
\end{aligned} \tag{3}$$

We want to estimate the number of nonzero terms in (3). Note that since E is fixed, if we choose values for $e^{(2)},\dots,e^{(r)}$, then there is at most one value of $e^{(1)}$ for which $a_{1e^{(1)}}\cdots a_{re^{(r)}}$ is a term in (3). Hence the number of nonzero terms in (3) is at most the number of ways to choose $e^{(2)},\dots,e^{(r)}$ such that $a_{2e^{(2)}},\dots,a_{re^{(r)}}$ are all nonzero. Applying (1) to each of $f_2,\dots,f_r$, we estimate

$$\prod_{i=2}^{r}(\#\text{ of nonzero } a_{ie}\text{'s}) \le \min\left\{\prod_{i=2}^{r}2^{\deg f_i},\ \prod_{i=2}^{r}(2\deg f_i)^n\right\}. \qquad (4)$$

To recapitulate, (4) gives an upper bound for the number of nonzero terms in the inner sums $\sum a_{1e^{(1)}}\cdots a_{re^{(r)}}$ appearing in (2). So if we let

$$N_v = \min\left\{\prod_{i=2}^{r}2_v^{\deg f_i},\ \prod_{i=2}^{r}(2\deg f_i)_v^n\right\},$$

then we have

$$\begin{aligned}\left|\prod_{i=1}^{r}f_i\right|_v &= \sup_E\left|\sum_{e^{(1)}+\cdots+e^{(r)}=E} a_{1e^{(1)}}\cdots a_{re^{(r)}}\right|_v\\ &\le N_v \sup_{e^{(1)},\dots,e^{(r)}}|a_{1e^{(1)}}\cdots a_{re^{(r)}}|_v\\ &\qquad\qquad \text{from (4) and the triangle inequality}\\ &= N_v\prod_{i=1}^{r}\sup_e|a_{ie}|_v\\ &= N_v\prod_{i=1}^{r}|f_i|_v.\end{aligned}$$

(b) We retain the notation from the proof of (a). Thus

$$\sum_{i=1}^{r}f_i=\sum_{i=1}^{r}\sum_e a_{ie}X^e=\sum_e\Bigl(\sum_{i=1}^{r}a_{ie}\Bigr)X^e,$$

so

$$\left|\sum_{i=1}^{r}f_i\right|_v=\sup_e\left|\sum_{i=1}^{r}a_{ie}\right|_v\le\sup_e\sup_i r_v|a_{ie}|_v=r_v\sup_i|f_i|_v.$$

(c) Again writing $f(X)=\sum_e a_eX^e$, every coefficient of $\partial f/\partial X_j$ has the form ma_e for some positive integer $m\le\deg f$ and some multi-index e. Hence

$$\left|\frac{\partial f}{\partial X_j}\right|_v\le\sup_e\sup_{m\le\deg f}|ma_e|_v=(\deg f)_v|f|_v.$$

(d) Continuing with the same notation as above, we have

$$\begin{aligned}
|f(b)|_v &= \Big|\sum_e a_e b^e\Big|_v \\
&= \Big|\sum_{e_1=0}^{d_1}\cdots\sum_{e_n=0}^{d_n} a_e b_1^{e_1}\cdots b_n^{e_n}\Big|_v \\
&\le \prod_{j=1}^n (d_j+1)_v \cdot \sup_e |a_e|_v \cdot \prod_{j=1}^n \sum_{0\le e_j\le d_j} |b_j|_v^{e_j} \\
&\le \prod_{j=1}^n (d_j+1)_v \cdot |f_v| \cdot \prod_{j=1}^n \max\{1,|b_j|_v\}^{d_j} \\
&\le \prod_{j=1}^n (d_j+1)_v \cdot |f_v| \cdot \max\{1,|b|_v\}^{\deg f}.
\end{aligned}$$

Combining this with the elementary estimate

$$\prod_{j=1}^n (d_j+1) \le \min\Big\{\prod_{j=1}^n 2_v^{d_j}, \prod_{j=1}^n (2\deg f)\Big\} \le \min\Big\{2^{\deg f}, (2\deg f)^n\Big\}$$

completes the proof of (d).
(e) Again we write $f = \sum a_e X^e$ and compute

$$\begin{aligned}
|f_b(X)|_v &= \Big|\sum_e a_e (X+b)^e\Big|_v \\
&= \Big|\sum_e a_e \Big(\sum_{i_1=0}^{e_1}\binom{e_1}{i_1} b_1^{e_1-i_1} X_1^{i_1}\Big)\cdots\Big(\sum_{i_n=0}^{e_n}\binom{e_n}{i_n} b_n^{e_n-i_n} X_n^{i_n}\Big)\Big|_v \\
&= \Big|\sum_{i_1=0}^{d_1}\cdots\sum_{i_n=0}^{d_n}\Big(\sum_{\substack{e_1,\dots,e_n\\ i_j\le e_j\le d_j}} a_e\binom{e_1}{i_1}\cdots\binom{e_n}{i_n} \\
&\qquad\qquad \times b_1^{e_1-i_1}\cdots b_n^{e_n-i_n}\Big) X_1^{i_1}\cdots X_n^{i_n}\Big|_v \\
&= \max_{\substack{i_1,\dots,i_n\\ 0\le i_j\le d_j}} \Big|\sum_{\substack{e_1,\dots,e_n\\ i_j\le e_j\le d_j}} a_e\binom{e_1}{i_1}\cdots\binom{e_n}{i_n} b_1^{e_1-i_1}\cdots b_n^{e_n-i_n}\Big|_v.
\end{aligned}$$

Now, the number of terms in this last sum is at most

$$\prod_{j=1}^n (d_j+1) \le \prod_{j=1}^n 2^{d_j} = 2^d;$$

and similarly we can estimate the binomial coefficients by

$$\binom{e_1}{i_1}\cdots\binom{e_n}{i_n} \le 2^{e_1+\cdots+e_n} \le 2^{d_1+\cdots+d_n} = 2^d.$$

(Of course, for nonarchimedean absolute values, the binomial coefficients have absolute value less than or equal to 1.) Using these observations we obtain the desired estimate

$$\begin{aligned}\left|f_b(X)\right|_v &\le 2_v^{2d}\max_e |a_e|_v \max\{1,|b_1|_v\}^{d_1}\cdots\max\{1,|b_n|_v\}^{d_n}\\ &\le 2_v^{2d}|f|_v\max\{1,|b|_v\}^{\deg f}.\end{aligned}$$ □

Remark B.7.5. One may easily convert the bounds of Proposition B.7.4 into bounds for heights. For example, (a) clearly implies (keeping the notation from Proposition B.7.4)

$$H(f_1\cdots f_r) \le \min\left\{\prod_{i=2}^{r}(2\deg f_i)^n, 2^{\deg f_2+\cdots+\deg f_r}\right\} H(f_1)\cdots H(f_r),$$

and inequality (e) implies

$$H(f_b) \le 4^{\deg f} H(f) H(b)^{\deg f}.$$

B.8. Local Height Functions

In this section and the next we will discuss, mainly without proof, the decomposition of height functions h_D into sums of local heights $\lambda_{D,v}$, one local height for each absolute value v of the field k. These decompositions are often essential for understanding the finer structure of height functions, but they will not be used in this book except at the very end (Part F), when we discuss various further results and open problems.

Let D be a divisor on a (smooth) variety V defined over k. For notational convenience, in this section we will write

$$V_D = V \smallsetminus \operatorname{supp}(D)$$

for the complement of the support of D. We would like to associate to each place $v \in M_k$ a function

$$\lambda_{D,v} : V_D(k_v) \longrightarrow \mathbb{R}$$

so that the sum

$$\sum_{v \in M_k} \lambda_{D,v}$$

is a Weil height function h_D for all points in $V_D(k)$. Further, the local height functions should be additive in D; and if D is a prime divisor (i.e., D is a irreducible subvariety of V of codimension 1), then $\lambda_{D,v}$ should be geometric in the following intuitive sense:

$$[\textit{Intuition}] \qquad \lambda_{D,v}(P) = -\log(v\text{-adic distance from } P \text{ to } D).$$

Thus as P gets v-adically close to D, the local height function becomes (logarithmically) larger. For a fixed point $P \in V(k)$, we also want the sum $\sum_v \lambda_{D,v}(P)$ to exist without worrying about convergence problems, so we will require that the local height $\lambda_{D,v}(P)$ vanish for all but finitely many $v \in M_k$.

To make all of this precise, we set some definitions. We define an M_k-*constant* to be a map

$$\gamma : M_k \longrightarrow \mathbb{R}$$

with the property that $\gamma_v = 0$ for all but finitely many $v \in M_k$. We say that a real-valued function ϕ on a subset Y of $V(k) \times M_k$ is M_k-*bounded* if there is an M_k constant γ such that

$$\left|\phi(P,v)\right| \le \gamma_v \quad \text{for all } (P,v) \in Y.$$

In particular, if $P \in V(k)$ is fixed, then $\phi(P,v) = 0$ for all but finitely many $v \in M_k$. When comparing functions, we will write $O_v(1)$ for an M_k-bounded function. Finally, we say that a subset Y of $V(k) \times M_k$ is *affine* M_k-*bounded* if there is an affine open subset $V_0 \subset V$ with affine coordinates $x_1, \dots, x_n$ such that $Y \subset V_0 \times M_k$ and such that the function

$$V_0(k) \times M_k \longrightarrow \mathbb{R}, \qquad P \longmapsto \max_{1 \le i \le n} \left|x_i(P)\right|_v,$$

is M_k-bounded on Y; and we say the set Y is M_k-*bounded* if it is a finite union of affine M_k-bounded sets.

For a given divisor D, we can think of λ_D as giving a family of functions, one for each $v \in M_k$. Equivalently, λ_D is a function on the disjoint union

$$\lambda_D : \coprod_{v \in M_k} V_D(k_v) \longrightarrow \mathbb{R},$$

where $\lambda_{D,v}$ is the restriction of λ_D to the set of v-adic points $V_D(k_v)$. Notice that there is a natural embedding of the set of ordered pairs

$$V_D(k) \times M_k$$

into the above disjoint union that takes an ordered pair (P,v) and identifies it with the point $P \in V(k_v)$. We will make use of this identification without further comment.

Theorem B.8.1. (Local height machine) *Let V/k be a smooth projective variety. For each $D \in \operatorname{Div}(V)$ it is possible to assign a function*

$$\lambda_D : \coprod_{v \in M_k} V_D(k_v) \longrightarrow \mathbb{R},$$

called the local height function with respect to *D, such that the following properties hold:*
(a) (Normalization) *Let $f \in k(V)^*$ be a rational function on V, and let $D = \operatorname{div}(f)$ be the divisor of f. Then the difference*

$$\lambda_{D,v}(P) - v(f(P))$$

is an M_k-bounded function on every M_k-bounded subset of $V_D(k) \times M_k$.
(b) (Additivity) *For all $D_1, D_2 \in \operatorname{Div}(V)$,*

$$\lambda_{D_1+D_2,v} = \lambda_{D_1,v} + \lambda_{D_2,v} + O_v(1).$$

(c) (Functoriality) *Let $\phi : V \to W$ be a morphism of smooth varieties. Then*

$$\lambda_{\phi^* D,v} = \lambda_{D,v} \circ \phi + O_v(1).$$

(d) (Positivity) *Let $D \geq 0$ be an effective divisor. Then*

$$\lambda_{D,v} \geq O_v(1).$$

(e) (Local/Global Property) *Let $D \in \operatorname{Div}(V)$, and let h_D be a Weil height attached to D. Then*

$$h_D(P) = \sum_{v \in M_k} d_v \lambda_{D,v}(P) + O(1) \quad \text{for all } P \in V_D(k),$$

where $d_v = [k_v : \mathbb{Q}_v]/[k : \mathbb{Q}]$.

PROOF. Let $D \in \operatorname{Div}(V)$. A candidate function for λ_D can be constructed as follows. Choose effective divisors $E_1, \ldots, E_n$ and $F_1, \ldots, F_m$ with the following properties:

$$\bigcap_{i=1}^{n} \operatorname{supp} E_i = \emptyset, \quad \bigcap_{j=1}^{m} \operatorname{supp} F_j = \emptyset, \quad \text{and} \quad D + E_i \sim F_j \quad \text{for all } i, j.$$

(It is always possible to find such divisors; see, for example, Lang [6, Chapter 10, Lemma 3.4].) Choose rational functions f_{ij} satisfying

$$\operatorname{div}(f_{ij}) = F_j - E_i - D.$$

Then a candidate for λ_D is the function

$$\lambda_{D,v}(P) = \max_{1\le j\le m}\min_{1\le i\le n}\log|f_{ij}(P)|_v.$$

Note that the poles of f_{ij} are on E_i and D, and the zeros are on F_j and D, so the fact that the E_i's (respectively the F_j's) have disjoint support means that $\lambda_{D,v}$ is well-defined off of the support of D. Further, if D is an effective divisor, then we see that $\lambda_{D,v}(P)$ tends to ∞ as P approaches D in the v-adic topology. This justifies our earlier intuitive description of the local height.

The key point now is to verify that the choice of divisors E_i, F_j and functions f_{ij} affects only λ_D by an M_k-bounded function. We refer the reader to Lang [6, Chapter 10] or Serre [3, Chapter 6.2] for the remainder of the proof of Theorem B.8.1 and for many further properties of local height functions. □

Remark B.8.2. The normalization property (B.8.1(a)) is most often used in the following way. Suppose that D_1 and D_2 are linearly equivalent divisors. Then $D_1 = D_2 + \operatorname{div}(f)$ for some rational function f, and (B.8.1(a,b)) implies that

$$\lambda_{D_1,v}(P) = \lambda_{D_2,v}(P) + v(f(P)) + O_v(1).$$

Remark B.8.3. It is possible to choose the local height functions on V consistently for field extensions. Thus for any finite extensions $L/K/k$ and any place $v \in M_K$, we have

$$\lambda_{D,v}(P) = \frac{1}{[L:K]}\sum_{w\in M_L,\, w|v}[L_w : K_v]\lambda_{D,w}(P) + O_v(1) \qquad \text{for all } P \in V_D(K_v).$$

Indeed, the construction of $\lambda_{D,v}$ as $\max_j \min_i \log|f_{ij}|_v$ gives this property immediately as soon as one knows that the construction is well-defined.

Example B.8.4. Let $V = \mathbb{P}^n$ and let D be a hypersurface defined by a homogeneous polynomial $Q(x_0, \ldots, x_n)$ of degree d. Then the function

$$\lambda_{D,v}(x) = \log \max_{0\le i\le n}\left|\frac{x_i^d}{Q(x_0,\ldots,x_n)}\right|_v$$

is a local height function associated to the divisor D.

B.9. Canonical Local Heights on Abelian Varieties

A Weil height function h_D associated to a divisor D is determined only up to a bounded function, and the important functoriality property $h_{\phi^*D} = h_D \circ \phi + O(1)$ holds only up to a bounded quantity. However, we have seen in Sections B.4 and B.5 that in certain cases it is possible to pick out particular Weil heights characterized by the property that functoriality holds exactly for certain divisors D and maps ϕ.

Similarly, the local height functions λ_D are determined only up to M_k-bounded functions, but just as for the Weil heights, it is sometimes possible to choose particular local height functions having particularly nice transformation properties. Thus the following theorem is related to Theorem B.8.1 in the same way that the canonical height theorem (B.4.1) is related to Weil's height machine (Theorem B.3.2).

Theorem 9.1. (Canonical local heights) *Let V/k be a smooth variety defined over a number field, let $D \in \mathrm{Div}(V)$, and let $\phi : V \to V$ be a morphism. Suppose that*

$$\phi^*D = \alpha D + \mathrm{div}(f)$$

for some number $\alpha > 1$ and some rational function $f \in k(V)^*$. Then there exists a local height function

$$\hat{\lambda}_{\phi,D} : \coprod_{v \in M_k} V_D(k_v) \longrightarrow \mathbb{R}$$

and an M_k-constant γ such that:

(i) $\hat{\lambda}_{\phi,D,v}(P) = \lambda_{D,v}(P) + O_v(1)$ *for all $P \in V_D(k_v)$.*

(ii) $\hat{\lambda}_{\phi,D,v}\big(\phi(P)\big) = \alpha\hat{\lambda}_{\phi,D,v}(P) + v\big(f(P)\big) + \gamma_v$ *for all $P \in V_D(k_v)$.*

Further, if we let $\hat{h}_{\phi,D}$ be the canonical height function defined in Theorem B.4.1 and let d_v be as in (B.8.1(e)), then there is a constant c such that

$$\hat{h}_{\phi,D}(P) = \sum_{v \in M_k} d_v \hat{\lambda}_{\phi,D,v}(P) + c \qquad \text{for all } P \in V_D(k).$$

PROOF. See Call–Silverman [1, Theorem 2.1]. □

Remark 9.2. With appropriate definitions, it is possible to generalize the theory of both global canonical heights (B.4.1) and local canonical heights (B.9.1) to include the extended divisor group $\mathrm{Div}(V) \otimes \mathbb{R}$. This is useful because the condition $\phi^*D \sim \alpha D$ may be true only for divisors in this extended group. See Silverman [6] for an example of a canonical height on a K3 surface V that uses an extended divisor D and a map $\phi : V \to V$ satisfying $\phi^*D \sim (7 + 4\sqrt{3})D$.

Just as in Section B.5, the general theory can be applied to the case of abelian varieties and the multiplication by m maps, since we know that $[m]^*D \sim m^2D$ (respectively $[m]^*D \sim mD$) if D is symmetric (respectively antisymmetric).

Theorem 9.3. (Néron [2]) *Let A/k be an abelian variety defined over a number field. For each divisor $D \in \operatorname{Div}(A)$ there is a local height function*

$$\hat{\lambda}_D : \coprod_{v \in M_k} A_D(k_v) \longrightarrow \mathbb{R},$$

called the canonical local height on A relative to D, satisfying the following conditions, where $\gamma_1, \gamma_2, \ldots$ denote M_k-constants:

(a) $\hat{\lambda}_{D,v} = \lambda_{D,v} + O_v(1)$.

(b) $\hat{\lambda}_{D_1+D_2,v} = \hat{\lambda}_{D_1,v} + \hat{\lambda}_{D_2,v} + \gamma_1(v)$.

(c) *If $D = \operatorname{div}(f)$, then $\hat{\lambda}_{D,v} = v \circ f + \gamma_2(v)$.*

(d) *If $\phi : B \to A$ is a homomorphism of abelian varieties, then $\hat{\lambda}_{\phi^*D,v} = \hat{\lambda}_{D,v} \circ \phi + \gamma_3(v)$.*

(e) *Let $Q \in A(k)$ and let $\tau_Q : A \to A$ be the translation-by-Q map. Then $\hat{\lambda}_{\tau_Q^*D,v} = \hat{\lambda}_{D,v} \circ \tau_Q + \gamma_4(v)$.*

(f) *Let $\hat{h}_D$ be the global canonical height function on A relative to D (B.5.6), and let d_v be as in (B.8.1(e)). Then there is a constant c such that*

$$\hat{h}_D(P) = \sum_{v \in M_k} d_v \hat{\lambda}_{D,v}(P) + c \qquad \text{for all } P \in A_D(k).$$

PROOF. See Lang [6, Chapter 11]. □

Remark 9.4. There are explicit formulas, due to Néron and Tate, for the canonical local heights on elliptic curves. See, for example, Silverman [2, Chapter VI].

Remark 9.5. Tate (unpublished) has given rapidly convergent series for the canonical local heights on elliptic curves over certain fields. These series have been generalized by Silverman [7] for elliptic curves and Call–Silverman [1] in general to give algorithms allowing the machine computation of canonical local and global heights reasonably efficiently, provided that the morphism ϕ is not too complicated.

B.10. Introduction to Arakelov Theory

We pursue briefly in this section the analogies between number fields and function fields. We first explain an explicit and geometric formulation of height theory over function fields and explain why this can be useful for Diophantine problems. We next show that this geometric formulation can be translated to the number field setting by (1) using schemes over Dedekind domains to extend varieties over number fields and (2) using complex analytic constructions such as Hermitian metrics or Green functions. This enables us to completely reformulate height theory (and proofs) in a more synthetic or geometric fashion. In the pursuit of further analogies, we are led to introduce special metrics, define extended divisors, etc., "à la Arakelov." For additional material on Arakelov theory, see, for example, Chinburg [1] or Lang [7].

We start with a smooth projective variety V defined over a function field $K = k(C)$, where C is a smooth projective curve over k. To simplify our exposition, we will assume that k is algebraically closed. We can construct a projective variety $\mathcal{V}$ over k with a morphism $\pi : \mathcal{V} \to C$ such that the generic fiber of π is isomorphic to V/K. We further assume that $\mathcal{V}$ is sufficiently smooth so that Weil divisors and Cartier divisors are the same. (For example, in characteristic 0 we could appeal to Hironaka's theorem on the resolution of singularities and assume that $\mathcal{V}$ is smooth.) A point $P \in V(K)$ induces a rational map $C \to \mathcal{V}$, and since C is smooth and $\mathcal{V}$ is projective, P will extend to a section (i.e., to a morphism) $\bar{P} : C \to \mathcal{V}$. Any divisor $D = \sum n_Y Y$ on V extends to a divisor $\bar{D}$ on $\mathcal{V}$ by taking the Zariski closure of each component and keeping the same multiplicities, say $\bar{D} := \sum n_Y \bar{Y}$. The divisor $\bar{D}$ is a Weil divisor, and hence is a Cartier divisor by hypothesis. Now observe that $\bar{P}^*(\bar{D})$ is well-defined as a divisor class on the curve C, and even as a divisor if we add the hypothesis that $P \notin \operatorname{supp}(D)$. We now define a function $h_{D,\mathcal{V}}$ on $V(K)$ by the formula

$$h_{D,\mathcal{V}}(P) = h_D(P) := \deg \bar{P}^*(\bar{D}) \qquad \text{for } P \in V(K).$$

Remark B.10.0. If P is a point defined over the algebraic closure of $K = k(C)$, say $P \in V(L)$ with L a finite extension of K, we can still define its "height" as follows. Fix a smooth projective curve C' and a covering $f : C' \to C$ such that $L = k(C')$ and such that the map f induces the inclusion $K \subset L$. The point P corresponds to a morphism $\bar{P} : C' \to \mathcal{V}$ as above, and we can define

$$h_{D,\mathcal{V}}(P) = h_D(P) := \frac{1}{[L:K]} \deg \bar{P}^*(\bar{D}).$$

One readily checks that this quantity is independent of the field L as long as $P \in V(L)$. Generally, the extension of height functions from $V(K)$ to $V(\bar{K})$ will be straightforward, so we will be content in this section to restrict attention to points in $V(K)$.

We now prove that our notation is consistent and that h_D defines a Weil height associated to the divisor D. We do this in several lemmas, where K will always denote a function field $k(C)$ as above.

Lemma B.10.1. *Let $Q = (f_0, \ldots, f_n) \in \mathbb{P}^n(K)$, let $\bar{Q}$ be the associated k-morphism $\bar{Q} : C \to \mathbb{P}^n$, and let H be a hyperplane (divisor class) in $\mathbb{P}^n$. Then*

$$\deg \bar{Q}^*(H) = \sum_{P \in C} \max_{0 \le i \le n} (-\operatorname{ord}_P(f_i)).$$

(Notice that the sum is nothing more than the height $h_K(Q)$ of the K-rational point in $\mathbb{P}^n$ for the usual collection of valuations on the function field K.)

Proof. Changing coordinates if necessary, we may assume that $\bar{Q}(C) \not\subset H_i$, where H_i is the hyperplane defined by $x_i = 0$. Let $D_i := \bar{Q}^*(H_i)$. Then

$$D_i - D_j = \bar{Q}^*(H_i) - \bar{Q}^*(H_j) = \operatorname{div}(f_i/f_j)$$

and

$$\operatorname{ord}_P D_i - \operatorname{ord}_P D_j = \operatorname{ord}_P f_i - \operatorname{ord}_P f_j,$$

and hence

$$\inf_i(\operatorname{ord}_P D_i) - \operatorname{ord}_P D_j = \inf_i(\operatorname{ord}_P f_i) - \operatorname{ord}_P f_j.$$

But since the D_i's are effective and the intersection of their supports is empty, we see that $\inf_i(\operatorname{ord}_P D_i) = 0$. Hence

$$-\inf_i(\operatorname{ord}_P f_i) = \operatorname{ord}_P D_j - \operatorname{ord}_P f_j.$$

Now summing over $P \in C$ and using the fact that $\sum_P \operatorname{ord}_P(f_j) = 0$ gives the desired result. □

Next we observe that if $Y \subset \mathcal{V}$ is an irreducible hypersurface on $\mathcal{V}$, then its image $\pi(Y)$ is either equal to all of C, or else it is equal to a single point. We say that D is a *vertical divisor* if π maps all of the components of D to points, and similarly we say that D is a *horizontal divisor* if π maps all of its components surjectively onto C. Clearly, any divisor can be written as the sum of a horizontal and a vertical divisor in a unique way. We also note that vertical divisors are characterized by the property that their restriction to the generic fiber of π is trivial.

Lemma B.10.2. *Let F be a vertical divisor on $\mathcal{V}$ (with respect to $\pi : \mathcal{V} \to C$). Then the map*

$$h_F : V(K) \to \mathbb{Z}, \qquad P \mapsto \deg \bar{P}^*(F),$$

takes finitely many values. In particular, h_F is a bounded function.

Proof. The proof is immediate once we note that if F is an irreducible component of a fiber of π, then $\deg \bar{P}^*(F) = 1$ if the section $\bar{P}$ meets

F, and $\deg \bar{P}^*(F) = 0$ otherwise. (We also observe that the extension $h_F : V(\bar{K}) \to \mathbb{Q}$ of h_F to $\bar{K}$ described in (B.10.0) is still a bounded function, but it may take infinitely many values.) □

We next use Lemma B.10.2 to show that changing the model $\mathcal{V}$ modifies the function h_D by only a bounded amount (in fact, by a function that takes finitely many values).

Lemma B.10.3. *Let $\pi : \mathcal{V} \to C$ and $\pi' : \mathcal{V}' \to C$ be two models for V/K. Then the difference $h_{D,\mathcal{V}} - h_{D,\mathcal{V}'}$ is bounded on $V(K)$.*

PROOF. We can find a third model $\mathcal{V}''$ that will dominate the other two, and hence we can reduce to the case where there is a birational morphism $f : \mathcal{V} \to \mathcal{V}'$ such that $\pi = \pi' \circ f$. If $P \in V(K)$ and $\bar{P}$ is the associated section from C to $\mathcal{V}$, then $\bar{P}' = f \circ \bar{P}$ is the section from C to $\mathcal{V}'$ corresponding to P'. (To see this, note that they coincide on a dense subset of C, hence are identical.) Now let D be a divisor on V, let $\bar{D}$ be the Zariski closure of D in $\mathcal{V}$, and let $\bar{D}'$ be the Zariski closure of D in $\mathcal{V}'$. Then the divisor $F := f^*(\bar{D}') - \bar{D}$ is trivial when restricted to the generic fiber, so F is a vertical divisor. Hence

$$h_{D,\mathcal{V}} - h_{D,\mathcal{V}'} = \deg \bar{P}^*(F)$$

is a bounded function by Lemma B.10.2. □

We are now ready to show that the geometrically defined h_D's give Weil heights for varieties defined over the function field K.

Theorem B.10.4. *For every variety V/K, fix a model $\pi : \mathcal{V} \to C$, and for every divisor D on V defined over K, define a function*

$$h_{D,\mathcal{V}} = h_D : V(K) \longrightarrow \mathbb{Z}, \qquad h_D(P) = \deg \bar{P}^* \bar{D}$$

as above. Then:
(a) $h_{D+D'} = h_D + h_{D'}$.
(b) *Let $f \in K(V)^*$ and $D = \operatorname{div}(f)$. Then $h_D = O(1)$.*
(c) *Let V/K and W/K be varieties, let $\phi : V \to W$ be a K-morphism, and let D be a divisor on W defined over K. Then $h_D \circ \phi = h_{\phi^*(D)} + O(1)$.*
(d) *Let $D \subset \mathbb{P}^n$ be a hyperplane defined over K, and let h be the usual Weil height on $\mathbb{P}^n(K)$. Then $h_D = h + O(1)$.*
(e) *If D is effective and $P \notin \operatorname{supp}(D)$, then $h_D(P) \geq 0$.*
In other words, the association $D \mapsto h_D$ from divisors to functions satisfies the axioms of a Weil height machine (cf. Theorem B.3.2).

PROOF. Property (a) is immediate by additivity of π^* and deg. To prove (b) we observe that $f \in K(V)$ will extend to a rational function $\bar{f}$ on $\mathcal{V}$

and that the restriction to the generic fiber of π of the two divisors $\operatorname{div}(\bar{f})$ and $\overline{(\operatorname{div}(f))}$ are the same; hence their difference is a vertical divisor, say

$$F := \operatorname{div}(\bar{f}) - \overline{(\operatorname{div}(f))}.$$

It follows from (B.10.2) that the function

$$h_{\operatorname{div}(f)}(P) = \deg \bar{P}^*(\overline{\operatorname{div}(f)}) = \deg \bar{P}^*(\operatorname{div}(\bar{f})) - \deg \bar{P}^*(F) = -\deg \bar{P}^*(F)$$

is bounded. To prove (c) we use the fact that we can choose models $\pi : \mathcal{V} \to C$ and $\pi' : \mathcal{W} \to C$ such that ϕ extends to a morphism $\bar{\phi}$ from $\mathcal{V}$ to $\mathcal{W}$ (see Exercises A.1.9 and A.9.5). Having done this, we see that $\bar{\phi}^*(\bar{D}) - \overline{\phi^*(D)}$ is trivial when restricted to the generic fiber of π; hence it is a vertical divisor F. Again by (B.10.2) and additivity, we conclude that $h_D \circ \phi - h_{\phi^*(D)}$ is bounded. Property (d) is just Lemma B.10.1 above. Finally, the effectivity of D implies that $\bar{D}$, and hence $\bar{P}^*(\bar{D})$, is also effective. Therefore $\bar{P}^*(\bar{D})$ has positive degree, which gives (e). □

We now want to build an analogue of the above construction when the function field is replaced by a number field. So we start with a number field K and a smooth projective variety V/K. We can construct a projective scheme $\pi : \mathcal{V} \to \operatorname{Spec}(R_K)$ with generic fiber V/K, and the fact that $\mathcal{V}$ is proper over $\operatorname{Spec}(R_K)$ implies that any rational point $P \in V(K)$ gives a section $\bar{P} : \operatorname{Spec}(R_K) \to \mathcal{V}$. Similarly, we can still define the closure of a divisor D on V to be its Zariski closure $\bar{D}$ in $\mathcal{V}$. If $\mathcal{V}$ is sufficiently smooth (e.g., if it is regular as an abstract scheme), then $\bar{P}^*(\bar{D})$ will give a well-defined divisor class on $\operatorname{Spec}(R_K)$, and indeed if the image of P does not lie in the support of D, then we get a well-defined divisor

$$\bar{P}^*(\bar{D}) = \sum_{v \in M_K^0} n_v[v].$$

(Recall (B.1.3) that the set of nonarchimedean absolute values on K is naturally identified with the set of prime ideals in R_K, hence our identification of M_K^0 with $\operatorname{Spec}(R_K)$. We will also write $\mathfrak{p}_v \in \operatorname{Spec}(R_K)$ for the prime ideal attached to the valuation $v \in M_K^0$.)

The question now is how to define the degree of $\bar{P}^*(\bar{D})$. As is clear from the function field construction, this degree should depend only on the divisor class of D, but the fact that $\operatorname{Spec}(R_K)$ is not "complete" means that intersection points may "move out to infinity." (To understand this analogy, try to construct a good intersection theory in the function field case for a family $\mathcal{V} \to \mathbb{A}^1$, and you will see the difficulty.)

The solution to this dilemma is to complete $\operatorname{Spec}(R_K)$ by allowing divisors that are supported at every place of K, not just at the nonarchimedean places. This leads to the following definition.

Definition. A *compactified (or Arakelov) divisor* on $\operatorname{Spec}(R_K)$ is a formal sum

$$E := \sum_{v\in M_K} m_v[v] \quad \text{with} \quad m_v \in \begin{cases} \mathbb{Z} & \text{if } v \in M_K^0, \\ \mathbb{R} & \text{if } v \in M_K^\infty. \end{cases}$$

A *principal compactified divisor* is a divisor of the form

$$\operatorname{div}(a) := \sum_{v\in M_K^0} \operatorname{ord}_v(a)[v] + \sum_{v\in M_K^\infty} -\log|a|_v[v]$$

for some $a \in K^*$. The *degree* of a compactified divisor $E = \sum m_v[v]$ is defined to be

$$\deg E := \sum_{v\in M_K^0} m_v \log N\mathfrak{p}_v - \sum_{v\in M_K^\infty} m_v[K_v : \mathbb{R}].$$

Observe that the product formula (Theorem B.1.2) says exactly that the degree of a principal compactified divisor is zero.

The idea now is to complete the divisor $\bar{P}^*(\bar{D}) = \sum_{v\in M_K^0} m_v[v]$ by adding to it a finite sum $\sum_{v\in M_K^\infty} m_v(D,P)[v]$ that takes account of the places "at infinity." One approach is to use *Green functions* (also called *Néron functions* in this context). For our purposes a Green function attached to a divisor D is simply a continuous function

$$G(D,\,\cdot\,) : V_D(\mathbb{C}) \longrightarrow \mathbb{R}$$

with a logarithmic pole along D. (Recall that $V_D := V \setminus \operatorname{supp}(D)$.) This last condition means that if U is an open subset of V and if $f = 0$ is a local equation for D on U, then the function

$$G(D,P) + \log|f(P)|,$$

defined a priori only on $U_D(\mathbb{C})$, extends to a continuous function on $U(\mathbb{C})$.

A variant of this point of view is furnished by the notion of a line bundle equipped with a norm or a metric. Note that a line bundle L on $\operatorname{Spec}(R_K)$ is simply a rank-one projective R_K-module. We create a metrized line bundle by adding metrics to the archimedean completions of L as explained in the following definition.

Definition. A *metrized line bundle on* $\operatorname{Spec}(R_K)$ is a rank-one projective R_K-module L together with a collection of (nontrivial) norms $\{|\cdot|_v\}_{v\in M_K^\infty}$ such that $|\cdot|_v$ is a norm on the K_v vector space $L_v := L \otimes K_v$ that is compatible with the norm on K_v.

The *Arakelov degree* of a metrized line bundle $(L, |\cdot|_v)$ is defined by picking any nonzero element $\ell \in L$ and setting

$$\deg_{\mathrm{Ar}}(L, |\cdot|_v) := \log\#(L/\ell R_K) - \sum_{v\in M_K^\infty} [K_v : \mathbb{R}] \log|\ell|_v.$$

As usual, the product formula (B.1.2) tells us that the degree is independent of the choice of ℓ.

Definition. Let V be a complex variety, and let $p : L \to V$ be a line bundle over V. A *metric on* L is a collection of norms $\{|\cdot|_x\}$, one for each $x \in V(\mathbb{C})$, such that $|\cdot|_x$ is a norm on the $\mathbb{C}$-vector space L_x and such that the metrics vary continuously as x varies. This last statement means that if $U \subset V$ is an open subset of V and if

$$\phi : U \times \mathbb{C} \xrightarrow{\sim} p^{-1}(U)$$

is a trivialization of L over U, then the function

$$U \times \mathbb{C} \longrightarrow \mathbb{R}, \qquad (x,c) \longmapsto \left|\phi(x,c)\right|_x,$$

is continuous on $U \times \mathbb{C}$.

Let $s : V \to L$ be a section to the line bundle L. We define the norm of s at the point $x \in V$ to be $|s(x)|_x$. Thus the norm of a section s is a continuous map

$$|s| : V \longrightarrow \mathbb{R}.$$

Now let V/K be an algebraic variety defined over a number field K. To each archimedean place $v \in M_K^\infty$ we associate a complex variety, denoted by V_v, by extending scalars to $\bar{K}_v \cong \mathbb{C}$.

Definition. Let K be a number field, let V/K be a smooth projective variety, let L be a line bundle on V defined over K, and let $P \in V(K)$. Choose a model $\mathcal{V} \to \operatorname{Spec}(R_K)$ of V and an extension $\mathcal{L}$ of L to $\mathcal{V}$. Also choose metrics $|\cdot|_v$ on L, one for each $v \in M_K^\infty$. Then the pullback $\bar{P}^*(\mathcal{L}, |\cdot|_v)$ is a metrized line bundle on $\operatorname{Spec}(R_K)$, and the *metrized height (or degree) of* P relative to these choices is

$$h_{\mathcal{V},\mathcal{L},|\cdot|_v}(P) := \frac{1}{[K:\mathbb{Q}]} \deg_{Ar} \bar{P}^*(\mathcal{L}, |\cdot|_v).$$

A *metrized height function (associated to L)* is any function h_L of the form $h_{\mathcal{V},\mathcal{L},|\cdot|_v}$ for any model $\mathcal{V} \to \operatorname{Spec}(R_K)$ for V, any extension $\mathcal{L}$ of L to $\mathcal{V}$, and any choice of metrics $|\cdot|_v$ on L.

To illustrate these abstract definitions, we will compute the metrized height function associated to the line bundle $\mathcal{O}(1)$ and the Fubini–Study metric on $\mathbb{P}^n$.

Example B.10.5. Let $V = \mathbb{P}^n$ and $L = \mathcal{O}(1)$. We choose $\mathcal{V} = \mathbb{P}^n_{R_K}$ and $\mathcal{L} = \mathcal{O}_{\mathcal{V}}(1)$. Let $a_0X_0 + \cdots + a_nX_n$ be a global section to L. The *Fubini–Study metric* $|\cdot|_{\mathrm{FS}}$ on L is defined by the formula

$$\left|(a_0X_0 + \cdots + a_nX_n)(P)\right|_{\mathrm{FS}}^2 := \left|\left(\frac{(a_0X_0 + \cdots + a_nX_n)^2}{X_0^2 + \cdots + X_n^2}\right)(P)\right|.$$

Then the associated metrized height function is

$$h_{\mathcal{V},\mathcal{L},|\cdot|_{\mathrm{FS}}}(P) = \sum_{v\in M_K^0} d_v \log \max_{0\le i\le n} |X_i(P)|_v + \sum_{v\in M_K^\infty} \frac{1}{2} d_v \log\left(\sum_{i=0}^{n} X_i^2(P)\right).$$

where as usual, $d_v = [K_v : \mathbb{Q}_v]/[K:\mathbb{Q}]$. We leave the verification of this formula as an exercise for the reader.

Notice that the Fubini–Study height in (B.10.5) differs by a bounded amount from the usual Weil height on $\mathbb{P}^n$. We will next show that in general, the choice of model $\mathcal{V}$ and metrics $|\cdot|_v$ affects the metrized height $h_{\mathcal{V},\mathcal{L},|\cdot|}$ by only a bounded amount. Following this, we will show that metrized heights are Weil heights, that is, they satisfy the properties of the Weil Height Machine (B.3.2).

Proposition B.10.6. *Let $\mathcal{V},\mathcal{L},|\cdot|_v$ and $\mathcal{V}',\mathcal{L}',|\cdot|'_v$ be two extensions with metrics of a variety V and a line bundle L on V as above. Then*

$$h_{\mathcal{V},\mathcal{L},|\cdot|_v}(P) = h_{\mathcal{V}',\mathcal{L}',|\cdot|'_v}(P) + O(1) \qquad \textit{for } P\in V(K).$$

PROOF. We consider first the case that $\mathcal{V} = \mathcal{V}'$. Let $\ell \ne 0$ be a section to L. Since $V(\mathbb{C})$ is compact, there exist constants $C_1, C_2 > 0$ such that $C_1 \le |\ell|_x/|\ell|'_x \le C_2$ for all $x \in V(\mathbb{C})$. This shows that the archimedean pieces of $h_{\mathcal{V},\mathcal{L},|\cdot|_v}$ and $h_{\mathcal{V},\mathcal{L}',|\cdot|'_v}$ differ by a bounded amount.

Next, since $\mathcal{L}'\otimes\mathcal{L}^{-1}$ is trivial on the generic fiber of $\mathcal{V}$, there exists a vertical divisor E such that $\mathcal{L}' = \mathcal{L}\otimes\mathcal{O}(E)$. The line bundle $\mathcal{O}(E)$ is trivial when restricted to the generic fiber; hence it may be equipped with the trivial metric, and then the function

$$P \longmapsto \deg_{\mathrm{Ar}} \bar{P}^*(\mathcal{O}(E), |\cdot|_v)$$

is bounded for $P \in V(K)$. This takes care of the nonarchimedean pieces, which proves that $h_{\mathcal{V},\mathcal{L},|\cdot|_v}$ and $h_{\mathcal{V},\mathcal{L}',|\cdot|'_v}$ differ by a bounded amount.

Finally, we consider the effect of choosing different models $\mathcal{V}$ and $\mathcal{V}'$. We may suppose that there is a birational morphism $f:\mathcal{V}'\to\mathcal{V}$ that is the identity on the generic fiber. We then choose $\mathcal{L}' := f^*\mathcal{L}$, and we take as a metric on $\mathcal{L}'$ the pullback of the metric on $\mathcal{L}$. If $P\in V(K)$, then the corresponding sections

$$\bar{P}' : \operatorname{Spec}(R_K) \to \mathcal{V}' \quad\text{and}\quad \bar{P} : \operatorname{Spec}(R_K)\to\mathcal{V}$$

are linked by $\bar{P} = f\circ\bar{P}'$. Therefore, $\bar{P}^*\mathcal{L} = (f\circ\bar{P}')^*\mathcal{L} = \bar{P}'^*\mathcal{L}'$, so in this case we get an equality $h_{\mathcal{V},\mathcal{L},|\cdot|_v}(P) = h_{\mathcal{V}',\mathcal{L}',|\cdot|'_v}(P)$. □

Proposition B.10.6 says that any two metrized height functions for L differ by a bounded amount. We now show that metrized height functions are Weil heights.

Theorem B.10.7. *Let K be a number field and V/K a variety as above.*
(a) Let L and L' be line bundles on V, and choose metrized heights h_L and $h_{L'}$ for L and L', respectively. Then $h_L + h_{L'}$ is a metrized height function associated to $L \otimes L'$.
(b) If L is the trivial bundle on V, then $h_L = 0$ is a metrized height function for L.
(c) Let $\phi : V \to W$ be a K-morphism between projective varieties, let L be a line bundle on W, and let h_L be a metrized height function for L. Then $h_L \circ \phi$ is a metrized height function for the line bundle $\phi^* L$ on V.
(d) The usual Weil height h on $\mathbb{P}^n(K)$ is a metrized height on $\mathbb{P}^n$ associated to the line bundle $O(1)$.
(e) There is a metrized height h_L for L such that $h_L(P) \geq 0$ for all P not in the base locus of L.

Remark B.10.8. By abuse of notation, people sometimes write Theorem B.10.7 as:
(a) $h_{L\otimes L'} = h_L + h_{L'}$.
(b) $h_0 = 0$.
(c) $h_L \circ \phi = h_{\phi^*(L)}$.
(d) $h_{O(1)} = h$.
(e) $h_L \geq 0$ off of the base locus of L.

Combining Proposition B.10.6 and Theorem B.10.7, we see that each of these formulas holds up to $O(1)$ for any choices of metrized heights. However, in order to get equality, Theorem B.10.7 says that the metrized heights must be carefully chosen.

PROOF (of Theorem B.10.7). (a) Fix a model for V, extensions $\mathcal{L}$ and $\mathcal{L}'$ for L and L', and metrics on L and L', corresponding to the choice of metrized heights h_L and $h_{L'}$. We take $\mathcal{L}\otimes\mathcal{L}'$ as our extension of $L\otimes L'$, and we take $|s\otimes s'|_v = |s|_v \cdot |s'|'_v$ as our family of metrics on $L\otimes L'$. Letting $h_{L\otimes L'}$ be the associated metrized height, the equality $h_{L\otimes L'} = h_L + h_{L'}$ is then clear from the definition of metrized height.
(b) The trivial metric on the trivial line bundle gives the zero function.
(c) Fix a model $\mathcal{W}$ for W, and extension $\mathcal{L}$ for L, and metrics $|\cdot|_v$ on L corresponding to the selected metrized height h_L. Choose a model $\mathcal{V}$ for V such that ϕ extends to a morphism $\bar{\phi} : \mathcal{V} \to \mathcal{W}$. (To do this, first choose any $\mathcal{V}$. Then ϕ extends to a rational map, and we can blow up to resolve the indeterminacy. See Hartshorne [1, II.7.17.3].) We take $\bar{\phi}^*\mathcal{L}$ as a model for $\phi^* L$ and the pullback metrics $|\phi^*(s)|_v = |s|_v$ as metrics on $\phi^* L$, and then the equality $h_{\phi^* L} = h_L \circ \phi$ is clear.
(d) Fix generators $X_0, \ldots, X_n$ for the space of global sections of the line bundle $O(1)$ on the scheme $\mathbb{P}^n_{\mathbb{Z}}$. That is, $X_0, \ldots, X_n$ are homogeneous coordinates on $\mathbb{P}^n_{\mathbb{Z}}$, or equivalently, $\mathbb{P}^n_{\mathbb{Z}} = \operatorname{Proj} \mathbb{Z}[X_0, \ldots, X_n]$. We define

metrics on $O(1)$ by the condition that for each section s to $O(1)$ and each point $P \in \mathbb{P}^n(K)$,

$$|s(P)|_v = \min_{\substack{0 \le i \le n \\ X_i(P) \ne 0}} \left|(f/X_i)(P)\right|_v.$$

With this choice of metrics, it is not hard to verify that the metrized height $h_{\mathbb{P}^n_{\mathbb{Z}}, O(1), |\cdot|_v}$ is the usual Weil height on $\mathbb{P}^n$. For further details, see Silverman [8, Proposition 7.2]. It is also instructive to compare with the metrized height on $O(1)$ for a different choice of metrics as described in (B.10.5).

(e) Take any model $\mathcal{V}$ for V, any extension $\mathcal{L}$ of L, and any set of metrics $|\cdot|_v$ on L. Let s be any nonzero section to L, and let $\bar{s}$ be its (unique) extension to $\mathcal{L}$. We can use the section $\bar{P}^*\bar{s}$ to $\bar{P}^*\mathcal{L}$ to compute

$$h_{\mathcal{V},\mathcal{L},|\cdot|_v}(P) = \frac{1}{[K:\mathbb{Q}]} \log \#\left(\bar{P}^*\mathcal{L}/\bar{s}(P)R_K\right) - \sum_{v \in M_K^\infty} d_v \log |s(P)|_v.$$

The first term on the right-hand side is clearly nonnegative. To deal with the sum over archimedean places, we define

$$\|s\|_\infty = \sup_{v \in M_K^\infty,\, P \in V(K_v)} |s(P)|_v.$$

Note that $\|s\|_\infty$ is finite, because $V(\mathbb{C})$ is compact and the various norms are continuous. Further, we need only a finite number of sections to define the base locus of L, so we have proven that there is a constant c (depending on all of our choices) such that

$$h_{\mathcal{V},\mathcal{L},|\cdot|_v}(P) \ge -c \quad \text{for all } P \text{ not in the base locus of } L.$$

Hence if we replace the original metrics by the equivalent metrics $|s|'_v := e^{-c}|s|_v$, we obtain a metrized height that is nonnegative off of the base locus of L. □

Exercises

B.1. Let $\phi : \mathbb{P}^n \longrightarrow \mathbb{P}^m$ be a rational map of degree d defined over $\bar{\mathbb{Q}}$. Write $\phi = [\phi_0, \ldots, \phi_m]$, where $\phi_i \in \bar{\mathbb{Q}}[X_0, \ldots, X_n]$ are homogeneous polynomials of degree d. Let A be the N-tuple consisting of all of the coefficients of all of the ϕ_i's, where we will consider A to be a point in $\mathbb{P}^N(\bar{\mathbb{Q}})$. Let $\operatorname{dom}(\phi) \subset \mathbb{P}^n(\bar{\mathbb{Q}})$ be the set on which ϕ is defined. Prove that

$$h(\phi(P)) \le dh(P) + h(\phi) + \log \binom{n+d}{n}$$

for all P in the domain of ϕ.

B.2. Let $\phi : \mathbb{P}^2 \to \mathbb{P}^2$ be the rational map $\phi(X,Y,Z) = (X^2, Y^2, XZ)$. Notice that ϕ is defined except at the point $(0,0,1)$.
(a) Let $P \in \mathbb{P}^2(\mathbb{Q})$ and choose homogeneous coordinates $P = (x,y,z)$ with $x,y,z \in \mathbb{Z}$ and $\gcd(x,y,z) = 1$. Prove that

$$h(\phi(P)) = \log \max\{|x^2|, |y^2|, |xz|\} - \log(\gcd(x, y^2)).$$

(b) Use (a) to show that there is no value of c such that the inequality $h(\phi(P)) \geq 2h(P) - c$ holds for all P.
(c) More generally, prove that the set

$$\left\{ \frac{h(\phi(P))}{h(P)} \;\middle|\; P \in \mathbb{P}^2(\mathbb{Q}) \text{ and } h(P) \neq 0 \right\}$$

is dense in the interval $[1,2]$.

B.3. Let V/k be a smooth variety defined over a number field. Let $D, E \in \operatorname{Div}(V)$ be divisors with D ample. Prove that there are constants c_1, c_2, depending on D and E, such that

$$|h_{V,E}(P)| \leq c_1 h_{V,D}(P) + c_2 \qquad \text{for all } P \in V(\bar{k}).$$

B.4. For any algebraic point $P \in \mathbb{P}^n(\bar{\mathbb{Q}})$, let d_P be the degree over $\mathbb{Q}$ of the field $\mathbb{Q}(P)$ generated by the coordinates of P, and let D_P be the absolute discriminant of $\mathbb{Q}(P)$ over $\mathbb{Q}$.
(a) Prove that

$$h(P) \geq \frac{1}{2d_P - 2}\left(\frac{1}{d_P} \log D_P - \log d_P\right).$$

(b) Show that the inequality in (a) is essentially best possible by taking $P = [a^{1/p}, 1]$ for an appropriate integer a and large prime number p.

B.5. Let V/k be a variety defined over a number field, let $\phi, \psi : V \to V$ be morphisms, and let $D \in \operatorname{Div}(V)$ be a divisor. Suppose that $\phi^* D \sim \alpha D$ and $\psi^* D \sim \beta D$ with $\alpha, \beta > 1$.
(a) If ϕ and ψ commute, prove that the associated canonical heights $h_{V,\phi,D}$ and $h_{V,\psi,D}$ are equal. That is, prove that

$$\phi \circ \psi = \psi \circ \phi \Longrightarrow h_{V,\phi,D}(P) = h_{V,\psi,D}(P) \quad \text{for all } P \in V(\bar{k}).$$

(b) Give an example to show that if ϕ and ψ do not commute, then the associated canonical heights need not be equal. (*Hint.* Use (B.4.2).)
(c) Let $V = \mathbb{P}^1$ and $D = (\infty)$. Prove that $\phi \circ \psi = \psi \circ \phi$ if and only if $h_{V,\phi,D} = h_{V,\psi,D}$. To what extent does this converse implication generalize to other varieties?

B.6. Let $a \in \mathbb{Z}$ be a nonzero square-free integer, and let $\phi : \mathbb{P}^1 \to \mathbb{P}^1$ be the map $\phi(x,y) = (2xy, x^2 + ay^2)$. Then $\phi^*(0,1) = (0,1) + (1,0) \sim 2(0,1)$, so there is a canonical height associated to ϕ and the divisor $D = (0,1)$. Find an explicit formula for this canonical height on $\mathbb{P}^1(\mathbb{Q})$. (*Hint.* This is one of the few rational maps on $\mathbb{P}^1$ for which it is possible to find a simple closed formula for the iterates ϕ^n.)

B.7. Let k be a number field and let E/k be an elliptic curve, say given by an equation

$$E : Y^2Z = X^3 + AXZ^2 + BZ^3 \qquad \text{with } A, B \in k,\ 4A^3 + 27B^2 \neq 0.$$

Also, let $x : E \to \mathbb{P}^1$ be the projection $x : (X, Y, Z) \mapsto (X, Z)$.
(a) For each integer $m \geq 1$, prove that there is a rational map $\phi_m : \mathbb{P}^1 \to \mathbb{P}^1$ of degree m^2 such that the following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{[m]} & E \\ \big\downarrow{\scriptstyle x} & & \big\downarrow{\scriptstyle x} \\ \mathbb{P}^1 & \xrightarrow{\phi_m} & \mathbb{P}^1 \end{array}$$

(b) Let $\hat{h}_{E,(0)}$ be the canonical height on E with respect to the divisor (0), and let $\hat{h}_{\mathbb{P}^1,\phi_m}$ be the canonical height on $\mathbb{P}^1$ with respect to the map ϕ_m and any divisor (p_0) on $\mathbb{P}^1$. Prove that

$$\hat{h}_{E,(0)}(P) = \hat{h}_{\mathbb{P}^1,\phi_m}(x(P)) \qquad \text{for all } P \in E(\bar{k}).$$

B.8. (a) Let $q : \mathbb{R}^r \to \mathbb{R}$ be a quadratic form. Prove that there is a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_r\}$ for $\mathbb{R}^r$ such that relative to this basis, q has the form

$$q\left(\sum_{i=1}^{r} x_i \mathbf{e}_i\right) = \sum_{i=1}^{s} x_i^2 - \sum_{i=1}^{t} x_{s+i}^2.$$

Prove that the integers s and t are uniquely determined by q. (This result was used in the proof of Corollary B.5.4.1.)
(b) Let Λ be a lattice in $\mathbb{R}^r$, let F be a fundamental domain for $\mathbb{R}^r/\Lambda$, and let $U \subset \mathbb{R}^r$ be a symmetric convex set. Prove that $\#(U \cap \Lambda) \geq 2^{-r} \operatorname{vol}(U)/\operatorname{vol}(F)$. (This generalizes Proposition B.5.4.)

B.9. Let A and B be abelian groups with B uniquely 2-divisible, let $h : A \to B$ be a quadratic function, and let

$$\langle \cdot, \cdot \rangle_h : A \times A \longrightarrow B, \qquad \langle P, Q \rangle_h = \frac{1}{2}\bigl(h(P+Q) - h(P) - h(Q) + h(0)\bigr),$$

be the associated symmetric bilinear pairing.
(a) Prove that the map

$$q : A \longrightarrow B, \qquad q(P) = \frac{1}{2}\bigl(h(P) + h(-P) - 2h(0)\bigr),$$

is a quadratic form on A and satisfies $q(P) = \langle P, P \rangle_h$.
(b) Prove that the map

$$\ell : A \longrightarrow B, \qquad \ell(P) = \frac{1}{2}\bigl(h(P) - h(-P)\bigr),$$

is a linear form on A.
(c) Let q and ℓ be as in (a) and (b), and let $b = h(0)$. Prove that $h = q + \ell + b$. Further, prove that this is the unique representation of h as the sum of a quadratic form, a linear form, and a constant.

B.10. Let V be a real vector space of dimension r, let $q : V \to \mathbb{R}$ be a positive definite quadratic form on V, and let $\Lambda \subset V$ be a lattice. Also, let λ be the first minimum of q on Λ, that is, $\lambda = \min\{q(x) \mid x \in \Lambda,\ x \neq 0\}$. Prove that

$$\#\{x \in \Lambda \mid q(x) \leq T\} \leq \left(\sqrt{4T/\lambda} + 1\right)^r.$$

(*Hint.* Find a value of N such that the given set maps injectively into $\Lambda/N\Lambda$.)

B.11. Let V/k be a smooth variety of dimension at least 1 defined over a number field k, and let $D \in \mathrm{Div}(V)$. Prove that the following two statements are equivalent:

(a) $|h_{V,D}(P)|$ is bounded for all $P \in V(\bar{k})$.

(b) D has finite order in $\mathrm{Pic}(V)$ (i.e., there is an integer $n \geq 1$ such that nD is linearly equivalent to 0).

B.12. Let C/k be a curve of genus $g \geq 2$, let K_C be a canonical divisor on C, and let $f : C \to J = \mathrm{Jac}(C)$ be the map defined by

$$f : C \longrightarrow J, \qquad P \longmapsto \mathrm{Cl}((2g-2)(P) - K_C).$$

(Note that $\deg K_C = 2g - 2$.)

(a) Prove that the map f is at most $(2g-2)^{2g}$-to-1.

(b) Let $\Theta \in \mathrm{Div}(J)$ be the theta divisor, let $\langle\cdot,\cdot\rangle_\Theta$ be the canonical height pairing attached to Θ, and let $\|\cdot\|_\Theta$ be the associated norm. Prove that for all $P, Q \in C(\bar{k})$ with $P \neq Q$,

$$\|f(P) - f(Q)\|_\Theta^2 \geq \left(1 - \frac{1}{g}\right)\left(\|f(P)\|_\Theta^2 + \|f(Q)\|_\Theta^2\right).$$

(*Hint.* Use Exercise A.8.2(c).) Compare this with the gap principle (B.6.6).

B.13. This exercise gives an explicit version of Lemma B.6.7 that is often useful for counting points on varieties. Let V be a real vector space of dimension n, let $\|\cdot\|$ be a Euclidean metric on V, and let Λ be a lattice in V. Let $S \subset \Lambda$ be a subset of Λ, and suppose that there are constants $a, b > 0$ such that

$$\|x - y\|^2 \geq a\left(\|x\|^2 + \|y\|^2\right) - b \qquad \text{for all } x, y \in S \text{ with } x \neq y.$$

If $T_2 \geq T_1 \geq \sqrt{b/a}$, prove that

$$\#\{x \in S \mid T_1 \leq \|x\| \leq T_2\} \leq \left(\log_2 \frac{2T_2}{T_1}\right)\left(\frac{4}{\sqrt{a}} + 1\right)^n.$$

(*Hint.* Use Exercise B.10.)

B.14. Let V be a finite-dimensional real vector space, let $\langle\cdot,\cdot\rangle$ be a Euclidean inner product on V, let $\|\cdot\|$ be the associated norm, and let Λ be a lattice of rank r in V.

(a) Prove that there exists a basis $u_1, u_2, \dots, u_r$ for Λ and an absolute constant $c_1 > 0$ such that

$$\left\|\sum_{i=1}^{r} a_i u_i\right\|^2 \geq c_1^r \sum_{i=1}^{r} a_i^2 \|u_i\|^2 \qquad \text{for all } a_1, \dots, a_r \in \mathbb{R}.$$

(b) The volume of a fundamental domain for Λ is equal to the square root of the determinant $\det(\Lambda) = \det\bigl(\langle v_i, v_j\rangle\bigr)_{i,j}$, where $v_1, \dots, v_r$ is any basis for Λ. Prove that there is an absolute constant $c_2 > 0$ such that the basis in (a) also satisfies

$$\sqrt{\det(\Lambda)} \leq \|u_1\| \cdot \|u_2\| \cdots \|u_r\| \leq c_2^{r^2} \sqrt{\det(\Lambda)}.$$

(The left-hand inequality holds for any basis. It is called Hadamard's inequality.)

A basis satisfying (a) or (b) is sometimes called *quasi-orthogonal*, since the angles between the basis elements cannot be extremely small.

B.15. Let V/k be a projective variety, let D and E be ample divisors on V, and let H_D and H_E be Weil height functions associated to D and E, respectively. Prove that the two counting functions $N(V(k), H_D, T)$ and $N(V(k), H_E, T)$ have roughly the same order of growth by proving that

$$\lim_{T\to\infty} \frac{\log\log N(V(k), H_D, T)}{\log\log N(V(k), H_E, T)} = 1.$$

B.16. Verify Manin's conjecture (B.6.9(a)) for the following varieties:

(a) Projective space $V = \mathbb{P}^n$.

(b) An abelian variety $V = A$.

B.17. Let V be a smooth variety defined over a number field k, and let $f \in k(V)^*$ be a rational function on V. Write the divisor of f as $\operatorname{div}(f) = \sum n_D D$, where each D is an effective irreducible divisor. Prove that

$$v(f(x)) = \sum_D n_D \lambda_{D,v}(x) + O_v(1).$$

This is a version of Weil's decomposition theorem.

B.18. Prove the formula for the height $h_{X,\mathcal{L},|\cdot|_{\mathrm{FS}}}$ associated to the Fubini–Study metric described in Example B.10.5. Show that this height differs by a bounded amount from the usual height on $\mathbb{P}^n$.

B.19. The purpose of this exercise is to compute the integral that appeared in the proof of Proposition B.7.3, namely $I(\alpha) = \int_0^1 \log|\alpha - \exp(2\pi i t)|\,dt$.

(a) Show that $I(\alpha)$ is well-defined for all $\alpha \in \mathbb{C}$, that it is continuous as a function of α, and that $I(\alpha) = I(|\alpha|)$.

(b) Show that $I(\alpha) = I(\alpha^{-1}) + \log|\alpha|$ for all $\alpha \in \mathbb{C}^*$.

(c) If $|\alpha| > 1$, integrate the function $f(z) = \log(z-\alpha)/z$ around the unit circle and apply the residue theorem (from complex analysis) to prove that $I(\alpha) = \log|\alpha|$.

(d) Conclude that $I(\alpha) = \log\max(1, |\alpha|)$.

B.20. Let $\alpha_1, \ldots, \alpha_r$ be algebraic numbers. Prove the elementary height inequalities

$$H(\alpha_1 + \cdots + \alpha_r) \le r \prod_{i=1}^{r} H(\alpha_i) \qquad \text{and} \qquad H(\alpha_1 \cdots \alpha_r) \le \prod_{i=1}^{r} H(\alpha_i).$$

B.21. Proposition B.7.2 uses inhomogeneous heights for polynomials, while Gelfand's inequality (B.7.3) uses homogeneous heights.
(a) Show that Proposition B.7.2(b) is false if we use homogeneous heights for the polynomials. (*Hint.* Take a polynomial whose coefficients have a common factor.)
(b) Show that Gelfand's inequality (B.7.3) is false if we use inhomogeneous heights for the polynomials. (*Hint.* What happens to the inequality if f_1 and f_2 are replaced with αf_1 and $\alpha^{-1} f_2$?)

B.22. Let $\mathcal{P} = \{P_{ji}; 1 \le i \le n,\ 1 \le j \le r\}$ be a collection of polynomials in $\bar{\mathbb{Q}}[X_1, \ldots, X_m]$ and set $h(\mathcal{P})$ equal, as usual, to the height of the point whose coordinates are all of the coefficients of all of the P_{ij}'s.
(a) Prove that

$$h\left(\sum_{i=1}^{n} P_{1i} \cdots P_{ri}\right) \le r h(\mathcal{P}) + \log n.$$

(b) Assume that $r = n$, and let $\Delta = \det(P_{ji})_{1 \le i,j \le r}$. Prove that

$$h(\Delta) \le r\left(h(\mathcal{P}) + \log r\right).$$

B.23. Let E be an effective divisor on a variety V, and let B_E be its base locus. The following two examples show that the restriction $P \notin B_E$ in Theorem B.3.2(e) is necessary to obtain the inequality $h_{V,E}(P) \ge O(1)$.
(a) Let $V \subset \mathbb{P}^n \times \mathbb{P}^{n-1}$ be the blowup of $\mathbb{P}^n$ at the point $P_0 = (0, \ldots, 0, 1)$ described in Example A.1.2.6(f). Let $p : V \to \mathbb{P}^n$ be the blowing-up map, let $q : V \to \mathbb{P}^{n-1}$ be the other projection, let E be the exceptional divisor, and let $M = q^*(\text{hyperplane } y_0 = 0)$. Show that

$$p^*(\text{hyperplane } x_0 = 0) = E + M,$$

and conclude that

$$h_E = h \circ p - h \circ q + O(1).$$

In other words, prove that

$$h_E(P) = h(x_0, \ldots, x_n) - h(y_0, \ldots, y_{n-1}) + O(1)$$

for all points $P = ((x_0, \ldots, x_n)(y_0, \ldots, y_{n-1})) \in V(\bar{\mathbb{Q}})$. Conclude directly (i.e., without using Theorem B.3.2(e)) that

$$h_{V,E}(P) \ge O(1) \qquad \text{for all } P \notin E,$$
$$h_{V,E}(P) = -h(y_0, \ldots, y_{n-1}) + O(1) \quad \text{for all points } P \in E.$$

In particular, observe that $h_E(P)$ is not bounded below for $P \in E$.
(b) Let C be a smooth curve of genus g, and let Δ be the diagonal of the product $V = C \times C$. If $g = 0$ or $g = 1$, prove that

$$h_{V,\Delta}(P, Q) \ge O(1) \qquad \text{for all } (P, Q) \in V(\bar{\mathbb{Q}});$$

but if $g \ge 2$, prove that $h_{V,\Delta}(P, P)$ is not bounded below.

PART C

Rational Points on Abelian Varieties

Progress has been much more general than retrogression.
C. Darwin, *The Descent of Man*

Our principal goal here in Part C is to prove the Mordell–Weil theorem.

Theorem C.0.1. (Mordell–Weil) *Let A be an abelian variety defined over a number field k. Then the group $A(k)$ of k-rational points of A is finitely generated.*

In the special case that the abelian variety is a cubic curve in the projective plane, we may express the result in the pleasing form, "There exists a finite set of rational points such that all rational points may be obtained from them by the tangent and chord process" (as described in (A.4.4)). This was Mordell's original formulation. Weil, in his thesis, extended Mordell's theorem to arbitrary number fields and to abelian varieties of higher dimension. More precisely, Weil dealt with Jacobians of curves of higher genus, since he had not yet developed the theory of abelian varieties.

Theorem C.0.1 can be generalized to fields finitely generated over their prime field (see Lang [6, Chapter 6, Theorem 1]). Using elementary group theory and the structure of the kernel of multiplication by m (Theorem A.7.2.7), we may rephrase Theorem C.0.1 by saying that there are points $P_1, \dots, P_r$ such that

$$A(k) = A(k)_{\text{tors}} \oplus \mathbb{Z}P_1 \oplus \cdots \oplus \mathbb{Z}P_r.$$

The integer r is called the *rank* of the abelian variety A/k, and $A(k)$ is the *Mordell–Weil group* of A/k. Note that the torsion subgroup $A(k)_{\text{tors}}$ is a finite abelian group, so it can be written as

$$A(k)_{\text{tors}} \cong (\mathbb{Z}/m_1\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/m_s\mathbb{Z}),$$

where $m_1, \dots, m_s$ are integers satisfying $m_i | m_{i+1}$ and $s \leq 2 \dim A$.

In this introduction we will use height theory and a descent argument to show that the following "weak Mordell–Weil theorem" implies the stronger version given above.

Theorem C.0.2. *("Weak" Mordell–Weil) Let A be an abelian variety defined over a number field k, let $A(k)$ be the group of k-rational points of A, and let $m \geq 2$ be an integer. Then the group $A(k)/mA(k)$ is finite.*

PROOF (that Theorem C.0.2 implies Theorem C.0.1). We select a symmetric ample divisor on A and let $\hat{h}$ denote the associated Néron–Tate height on $A(k)$ (see Chapter B.5). Recall that $\hat{h}$ is a nonnegative quadratic form on $A(k)$ with the property that for all $C > 0$, the set

$$\{x \in A(k) \mid \hat{h}(x) \leq C\}$$

is finite. The following lemma axiomatizes this situation and completes the proof that (C.0.2) implies (C.0.1).

Lemma C.0.3. *(Descent lemma) Let G be an abelian group equipped with a quadratic form $q : G \to \mathbb{R}$ such that for all C, the set*

$$\{x \in G \mid q(x) \leq C\}$$

is finite. Assume further that for some integer $m \geq 2$, the group G/mG is finite. Then G is finitely generated.

More precisely, let $g_1, \ldots, g_s$ be a set of representatives for G/mG, and let $C_0 := \max_i q(g_i)$. Then G is generated by the finite set

$$\{x \in G \mid q(x) \leq C_0\}.$$

PROOF (of the descent lemma). We begin by observing that for all x in G, we have $q(x) \geq 0$, since otherwise we would have infinitely many points with $q(x)$ negative. To ease notation, we may therefore safely put

$$|x| := \sqrt{q(x)}, \qquad c_0 := \max_i |g_i|, \qquad \text{and} \qquad S = \big\{x \in G \mid |x| \leq c_0\big\}.$$

We will prove that the finite set S generates G.

Let $x_0 \in G$. If $x_0 \in S$, we are done. Otherwise, $|x_0| > c_0$, so we consider the image of x_0 in G/mG and choose a coset representative g_i for x_0. This means that $x_0 = g_i + mx_1$ for some $x_1 \in G$. We use the triangle inequality to compute

$$\begin{aligned} m|x_1| &= |x_0 - g_i| \\ &\leq |x_0| + |g_i| && \text{by the triangle inequality} \\ &< 2|x_0| && \text{since } |g_i| \leq c_0 < |x_0|. \end{aligned}$$

Since $m \geq 2$ by assumption, we find that $|x_1| < |x_0|$.

If $x_1 \in S$, then $x_0 = g_i + mx_1$ is in the subgroup generated by S, and we are done. Otherwise we can write $x_1 = g_j + mx_2$, and the same

computation reveals that $|x_2| < |x_1|$. Continuing in this fashion, we obtain a sequence of elements $x_0, x_1, x_2, \dots$ satisfying $|x_0| > |x_1| > |x_2| > \cdots$ and with the property that for each t, the initial element x_0 is a linear combination of x_t and $g_1, \dots, g_s$. Since G has only finitely many elements of bounded size, eventually the sequence terminates with an element that is in S, which completes the proof that x_0 is a linear combination of the elements of S. Therefore, S generates G. □

Remark C.0.4. (a) The term "descent" comes from Fermat's famous *descente infinie*. Indeed, the arguments are very similar. For example, Fermat's proof that integral solutions to $x^2 - dy^2 = 1$ are generated by one fundamental solution and his proof that the equation $x^4 + y^4 = z^2$ has no nontrivial integral solutions can be rewritten along the lines of the proof of (C.0.3). (See Knapp [1, Chapter IV] for details.) Roughly speaking, one builds a "smaller" solution from a given solution, and repeating the process eventually yields either a contradiction or a set of generating solutions.

(b) There is an obvious (but tedious) effective process to find all points of bounded height on an abelian variety. Hence if we could effectively find coset representatives for $A(k)/mA(k)$, we would be able to effectively compute generators of the group $A(k)$. Unfortunately, no such a process is known today. (We will comment on this further later in this chapter, see especially the remarks in Section 4).

(c) The proof of the weak Mordell–Weil theorem will give an effective bound for the order of $A(k)/mA(k)$, and hence will yield an effective bound for the rank of $A(k)$ (see Theorem C.1.9).

(d) It is not necessary to have a refined theory of heights in order to deduce the full Mordell–Weil theorem from the weak Mordell–Weil theorem. It suffices to have some fairly crude height inequalities. See Exercise C.1 for details.

(e) It clearly suffices to prove Theorem C.0.1 with the field k replaced by any finite extension of k, since a subgroup of a finitely generated group is again finitely generated. This reduction to a larger field is not as straightforward for Theorem C.0.2, but we will give a proof below (Lemma C.1.1).

We conclude this introduction with an outline of the proof of the weak Mordell–Weil theorem (C.0.2). Filling in the details of the proof will occupy much of the rest of Part C.

Let L be the field obtained by adjoining to k the coordinates of the point $Q \in A(\bar{k})$ satisfying $mQ \in A(k)$. Intuitively, these are the points obtained by "dividing" the points in $A(k)$ by m. Then a general argument using Galois theory, which works over any field, shows that $A(k)/mA(k)$ is finite if and only if the field L is a finite extension of k.

If we assume, as we may, that all of the m-torsion points of A are k-rational, then L/k will be an abelian extension of k of exponent m; that is, $\mathrm{Gal}(L/k)$ is abelian and every element of $\mathrm{Gal}(L/k)$ has order dividing m.

We now use the arithmetic fact that an abelian extension of exponent m is a finite extension if and only if it is unramified outside a finite set of primes. This can be proven using Kummer theory or Hermite's theorem.

This reduces the proof to studying the ramified primes in the extension L/k. We will show that L/k is unramified outside the places of bad reduction of A and the places dividing m. This will follow from the crucial fact that if A has good reduction at $\mathfrak{p}$, then reduction modulo $\mathfrak{p}$ is injective on torsion points of order prime to p, where p is the characteristic of $\mathfrak{p}$. (See Theorem C.1.4, proven in Section C.2.)

Notice that this injectivity provides an effective and efficient way to compute the torsion part of the Mordell–Weil group. The infinite part of $A(k)$ is much more difficult to compute. In fact, no algorithm (guaranteed to terminate) is known, although a reasonable one often works in practice. In Section C.3 we review some background material and prove the basic finiteness theorems of algebraic number theory. Section C.4 provides further details on the descent argument. In that section we rephrase the descent argument in terms of Galois cohomology and show explicitly how the problem of making descent effective is tied into the failure of the Hasse principle. Finally, in Section C.5 we give basic definitions and properties of group cohomology. Also note that an explicit version of some of the computations in this chapter are given for the Jacobian of hyperelliptic curves in the series of exercises C.18 and C.19.

C.1. The Weak Mordell–Weil Theorem

Recall that the multiplication-by-m map $[m] : A(\bar{k}) \to A(\bar{k})$ is surjective with finite kernel, denoted by A_m, and that A_m is isomorphic (as an abstract group) to $(\mathbb{Z}/m\mathbb{Z})^{2g}$. For each $x \in A(k)$, we select a point $y \in A(\bar{k})$ satisfying $[m](y) = x$, and then for each $\sigma \in \mathrm{Gal}(\bar{k}/k)$ we consider the point

$$\sigma(y) - y \in A_m.$$

Note that $\sigma(y) - y$ is in A_m, since

$$[m](\sigma(y) - y) = \sigma([m]y) - [m]y = \sigma(x) - x = 0.$$

We begin by using this construction to show that it suffices to prove Theorem C.0.2 with k replaced by a finite extension.

Lemma C.1.1. *For any finite extension k'/k, the kernel of the natural map $A(k)/mA(k) \to A(k')/mA(k')$ is finite.*

PROOF. Replacing k' by its Galois closure over k only makes the statement of the lemma stronger, so we may assume that k'/k is Galois with group G. The kernel of the map $A(k)/mA(k) \to A(k')/mA(k')$ is

$$B := \big(A(k) \cap mA(k')\big)/mA(k).$$

For any element $x \pmod{mA(k)}$ in B, we fix an element $y \in A(k')$ such that $[m](y) = x$, and then we define a map

$$f_x : G \longrightarrow A_m(k'), \qquad f_x(\sigma) = y^\sigma - y.$$

The map f_x is not a homomorphism in general, but in any case it is a map from G to A_m considered as sets. We have thus defined a map

$$B \longrightarrow \mathrm{SetMaps}(G, A_m), \qquad x \longmapsto f_x.$$

Since G and A_m are both finite sets, if we can show that this map is injective, we will have shown that B is finite.

Suppose that $f_x = f_{x'}$, and let y, y' be the points used to define $f_x, f_{x'}$. Then

$$f_x(\sigma) = \sigma(y) - y = \sigma(y') - y' = f_{x'}(\sigma) \qquad \text{for every } \sigma \in G,$$

so $\sigma(y - y') = y - y'$ for every $\sigma \in G$. This implies that $y - y' \in A(k)$, so

$$x - x' = [m]y - [m]y' = [m](y - y') \in mA(k).$$

This means that x and x' represent the same element in B, which completes the proof of Lemma C.1.1. (The ad hoc argument given here will be rephrased in Section C.3 in terms of cohomology. In particular, the map $f_x : G \to A_m$ is a 1-cocycle, and it is well-defined as a cohomology class.) □

For the rest of this section, we will make the following assumptions:

- A_m is contained in $A(k)$.
- μ_m (the mth roots of unity) are contained in k.

(Actually, the former assumption implies the latter using Weil's pairing; see Exercise A.7.8.) With these assumptions, we make the following definition.

Definition. For each $x \in A(k)$ and each $\sigma \in \mathrm{Gal}(\bar{k}/k)$, choose some $y \in A(\bar{k})$ satisfying $[m](y) = x$ and define

$$t(\sigma, x) := \sigma(y) - y.$$

We verify below (under our assumption $A_m \subset A(k)$) that the value of $t(\sigma, x)$ depends only on x, and not on the choice of y. The resulting map

$$t : \mathrm{Gal}(\bar{k}/k) \times A(k) \longrightarrow A_m$$

is called the *Kummer pairing on A*. Notice the analogy with the classical Kummer pairing,

$$\mathrm{Gal}(\bar{k}/k) \times k^* \longrightarrow \mu_m, \qquad (\sigma, a) \longmapsto \sigma(\alpha)/\alpha \quad (\text{where } \alpha = \sqrt[m]{a}).$$

The most important properties of the Kummer pairing t are given in the following proposition.

Proposition C.1.2. *Assume* $A_m \subset A(k)$.
(i) *The function* $t(\sigma, x) = \sigma(y) - y$ *described above is well-defined and gives a bilinear map* $t : \mathrm{Gal}(\bar{k}/k) \times A(k) \to A_m$.
(ii) *Let* L *be the extension of* k *obtained by adjoining to* k *the coordinates of all points* $y \in A(\bar{k})$ *satisfying* $[m](y) \in A(k)$. *Then* t *induces a nondegenerate pairing*

$$t : \mathrm{Gal}(L/k) \times A(k)/mA(k) \longrightarrow A_m.$$

In particular, $A(k)/mA(k)$ *is finite if and only if* L *is a finite extension of* k.

PROOF. We begin by checking that $t(\sigma, x)$ does not depend on the choice of y. So suppose that $[m](y') = [m](y) = x$. Then $y' - y \in A_m \subset A(k)$, so

$$\big(\sigma(y) - y\big) - \big(\sigma(y') - y'\big)\sigma(y - y') - (y - y') = 0.$$

Therefore, the value of $t(\sigma, x)$ depends only on σ and x, so t is well-defined.

The verification of bilinearity is similarly straightforward. We start with the first variable:

$$\begin{aligned} t(\sigma\sigma', x) &= y^{\sigma\sigma'} - y = (y^\sigma - y)^{\sigma'} + (y^{\sigma'} - y) \\ &= t(\sigma, x)^{\sigma'} + t(\sigma', x) = t(\sigma, x) + t(\sigma', x). \end{aligned}$$

Note that the last equality is true because $t(\sigma, x) \in A_m \subset A(k)$, so $t(\sigma, x)$ is invariant by Galois.

Next we compute

$$t(\sigma, x + x') = (y + y')^\sigma - (y + y') = (y^\sigma - y) + (y'^\sigma - y') = t(\sigma, x) + t(\sigma, x').$$

This completes the proof of the bilinearity of t.

Since A_m has exponent m, the left kernel certainly contains $mA(k)$. Now suppose that x is in the left kernel. This means that $\sigma(y) = y$ for all $\sigma \in \mathrm{Gal}(\bar{k}/k)$, and hence that $y \in A(k)$ and $x = [m](y) \in mA(k)$. Therefore, the left kernel is exactly $mA(k)$.

Next we observe that an element σ is in the right kernel if and only if for every $y \in A(\bar{k})$ satisfying $[m](y) \in A(k)$ we have $\sigma(y) = y$. From the definition of L, this is equivalent to saying that $\sigma \in \mathrm{Gal}(\bar{k}/L)$. Hence the right kernel is $\mathrm{Gal}(\bar{k}/L)$. Taking the quotient by the left and right kernels gives a nondegenerate pairing as stated in the theorem. □

We thus need to understand the ramification properties of field extensions of the form $k(y)/k$, where $[m](y) = x \in A(k)$. Under our assumption that $A_m \subset A(k)$, the extension $k(y)$ depends only on x and is independent of the choice of y, so to simplify notation we will denote such an extension by $k(\frac{1}{m}x)$.

Lemma C.1.3. *Assume $A_m \subset A(k)$ and $x \in A(k)$. Then the extension $k(\frac{1}{m}x)$ is Galois over k, and its Galois group is canonically isomorphic to a subgroup of A_m.*

PROOF. Fix a point $y \in A(\bar{k})$ with $[m](y) = x$. We have seen above that the Galois conjugates of y differ by elements of A_m, so the assumption that $A_m \subset A(k)$ implies that all of the Galois conjugates of y are already in $k(y)$. Thus $k(y)$ is Galois over k. Further, the map

$$\mathrm{Gal}(k(y)/k) \longrightarrow A_m, \qquad \sigma \longmapsto t(\sigma, x) = \sigma(y) - y,$$

is a group homomorphism (C.1.2(i)), and it is injective, since σ is determined by its action on y. Hence $\mathrm{Gal}(k(y)/k)$ is isomorphic to a subgroup of A_m. □

Classical Kummer theory (see below) now tells us that $k(\frac{1}{m}x)$ is obtained by adjoining to k some m^{th}-roots of elements of k. Up to now, we have not used any special properties of the field k, but to go further, we must use the arithmetic nature of k. We will need the following key result, whose proof occupies the next section.

Theorem C.1.4. *Let A be an abelian variety defined over a number field k, let v be a finite place of k at which A has good reduction, let $\tilde{k}$ be the residue field of v, and let p be the characteristic of $\tilde{k}$. Then for any $m \geq 1$ with $p \nmid m$, the reduction map*

$$A_m(k) \longrightarrow \tilde{A}(\tilde{k})$$

is injective. In other words, the reduction modulo v map is injective on the prime-to-p torsion subgroup of $A(k)$.

We observe that Theorem C.1.4 immediately implies that the torsion part of $A(k)$ is finite. Indeed, by choosing two places v and w of good reduction and of different characteristics, we obtain an injection

$$A(k)_{\text{tors}} \hookrightarrow \tilde{A}_v(\tilde{k}_v) \times \tilde{A}_w(\tilde{k}_w),$$

and the latter is clearly a finite group. This observation is often the easiest way to determine the torsion subgroup; see Exercises C.3, C.5, and C.6 for some explicit computations.

Proposition C.1.5. *Let $m \geq 1$ be an integer, and let S be the (finite) set of places of k at which A has bad reduction, together with the places that divide m. Then for all rational points $x \in A(k)$, the extension $k(\frac{1}{m}x)/k$ is unramified outside S.*

Hence the field L described in (C.1.2(ii)) is unramified outside S.

PROOF. Note that the crucial fact being proven in this proposition is that the set of possibly ramified places S can be chosen independently of the choice of the point $x \in A(k)$.

We choose a point $y \in A(\bar{k})$ satisfying $[m]y = x$ as usual, and to ease notation, we let $k' = k(y) = k(\frac{1}{m}x)$. Let v be finite a place of k not in S, and let w be any extension of v to k'. We consider the reduction modulo w map

$$A(k') \longrightarrow \tilde{A}_w(\tilde{k}'_w).$$

If $\sigma \in \mathrm{Gal}(k'/k)$ is in the decomposition group of w, we get by reduction an automorphism $\tilde{\sigma}$ in the Galois group $\mathrm{Gal}(\tilde{k}'_w/\tilde{k}_v)$ of the residue fields, and σ is in the inertia group for w if and only if $\tilde{\sigma} = 1$

So suppose that σ is in the inertia group for w. This means that $\tilde{\sigma}$ acts trivially on $\tilde{k}'_w$, so $\tilde{\sigma}$ acts trivially on $\tilde{A}_w(\tilde{k}'_w)$, and hence $\tilde{\sigma}(\tilde{y}) = \tilde{y}$. But this implies that

$$\widetilde{t(\sigma, y)} = \widetilde{\sigma(y)} - y = \tilde{\sigma}(\tilde{y}) - \tilde{y} = \tilde{0}$$

directly from the definition of the pairing t. Now Theorem C.1.4 tells us that the m-torsion of $A(k')$ injects into the reduction $\tilde{A}(\tilde{k}'_w)$, so the fact that the reduction $\widetilde{t(\sigma, y)}$ is zero lets us conclude that

$$t(\sigma, y) = 0.$$

In other words, $\sigma(y) = y$, so σ acts trivially on k', so $\sigma = 1$. This proves that the inertia group of w is trivial, which is equivalent to the assertion that k'/k is unramified at w. Since v was an arbitrary finite place not in S, and w was an arbitrary place of k' lying over v, this completes the proof that k' is unramified outside of S. □

It is now possible to quickly finish the proof of Theorem C.0.2 by using Proposition C.1.5 and the following fundamental result from algebraic number theory.

Proposition C.1.6. (Hermite) *Let k be a number field, let d be a postive integer, and let S be a finite set of places of k. Then there are only a finite number of extensions of k of degree less than d and unramified outside S.*

PROOF. This classical result is usually proven in two steps. First one shows that the discriminant of such an extension is bounded, and then that there exist only a finite number of extensions of a given degree and discriminant. See Theorem C.3.2 below, or Serre [4, Proposition 7.13], Samuel [1, Théorème 3 Chapitre 4.3] or Lang [9, Theorem 5, V.4]. □

Remark C.1.6.1. We observe that it is possible to avoid the use of Theorem C.1.4 by appealing to the more elementary Chevalley–Weil theorem. This theorem states that if $f : X \to Y$ is a finite unramified map, then there is a finite set of places S such that for any rational point $y \in Y(k)$, the field generated by $f^{-1}(y)$ is unramified outside of S. Hence the compositum of the fields $k(f^{-1}(y))$ is finite over k. (See Exercise C.7 or Lang [6,

Theorem 8.1].) However, we want to be more precise and to give a bound for the degree of the compositum L that we are considering. To do that, we will dig deeper into the structure of abelian extensions of exponent m. We know by Kummer theory (see Lang [2, VIII.8, Theorems 13 and 14], for example) that an abelian extension of k of exponent m corresponds to a subgroup B of k^* containing $(k^*)^m$. The correspondence is given by

$$B \longmapsto k\left(\sqrt[m]{B}\right) \qquad \text{and} \qquad L \longmapsto (L^*)^m \cap k^*.$$

We need to analyze ramification in such extensions.

Lemma C.1.7. *Assume that k contains a primitive m^{th}-th root of unity. Let $\alpha \in k^*$, let $K := k(\sqrt[m]{\alpha})$, and let v be a place of k not dividing m. Then the extension K/k is unramified at v if and only if* $\operatorname{ord}_v(\alpha) \equiv 0 \pmod{m}$.

PROOF. To ease notation, let $\omega = \sqrt[m]{\alpha}$. An easy calculation shows that the discriminant of ω (equivalently the discriminant of the order $R_k[\omega]$) is equal to $m^m\alpha^{m-1}$ (see Exercise C.8), so the discriminant of K/k divides $m^m\alpha^{m-1}$. Since $\operatorname{ord}_v(m) = 0$ by assumption, this shows that if $\operatorname{ord}_v(\alpha) = 0$, then K/k is unramified at v. It remains to consider the case that $\operatorname{ord}_v(\alpha) > 0$.

Suppose first that $\operatorname{ord}_v(\alpha) \equiv 0 \pmod{m}$, say $\operatorname{ord}_v(\alpha) = mt$. Let $\pi \in k$ be a uniformizer at v, i.e., $\operatorname{ord}_v(\pi) = 1$, and consider the element $\beta = \alpha\pi^{-mt}$. Then $K = k(\sqrt[m]{\alpha}) = k(\sqrt[m]{\beta})$ and $\operatorname{ord}_v(\beta) = 0$, so from our discussion above we conclude that K is unramified at v.

It remains to deal with the case that $\operatorname{ord}_v(\alpha) \not\equiv 0 \pmod{m}$. Let $r = \operatorname{ord}_v(\alpha)$, and let $\mathfrak{p}$ be the prime ideal of k corresponding to v. Then we can write $\alpha R_k = \mathfrak{p}^r\mathfrak{A}$ for some ideal $\mathfrak{A}$ relatively prime to $\mathfrak{p}$. Suppose that the ideal $\mathfrak{p}$ splits in K as a product $\mathfrak{p}R_K = \mathfrak{P}_1^{e_1}\cdots\mathfrak{P}_s^{e_s}$ of prime ideals. Then

$$\alpha R_K = \mathfrak{P}_1^{re_1}\cdots\mathfrak{P}_s^{re_s}\mathfrak{A}'$$

for an ideal $\mathfrak{A}'$ of R_K relatively prime to all of the $\mathfrak{P}_i$'s. The principal ideal αR_K is an mth power in K, since $\alpha = \omega^m$, so we conclude that $m | re_i$ for all $1 \le i \le s$. Since $m \nmid r$ by assumption, it follows that every $e_i \ge 2$, and hence K/k is ramified over v. □

Corollary C.1.8. *Let k be a number field, m an integer, and S a finite set of finite places of k. Assume that k contains a primitive m^{th}-root of unity, that S contains all places of k dividing m, and that the ring of S-integers $R_{k,S}$ is a principal ideal domain. (This last condition can always be achieved by enlarging S.) Let K be the maximal extension of k such that K/k is abelian of exponent m and is unramified outside S.*

(a) *The field K is equal to $k\big((R_{k,S}^*)^{1/m}\big)$. That is, K is the field obtained by adjoining to k all of the mth roots of all of the elements of $R_{k,S}^*$.*

(b) *The field K is a finite Galois extension of k with Galois group*

$$\mathrm{Gal}(K/k) \cong (\mathbb{Z}/m\mathbb{Z})^{1+r(S)},$$

where $r(S)$ is the rank of $R^*_{k,S}$. Further,

$$r(S) = r_1 + r_2 - 1 + |S|,$$

where r_1 (respectively r_2) is the number of real embeddings (respectively pairs of conjugate complex embeddings) of k into $\mathbb{C}$.

Proof. (Compare with Lang [9, Theorem 1, Chapter XI.2]). Let

$$K' = k\left(\sqrt[m]{R^*_{k,S}}\right),$$

and let K be as in the statement of the theorem. From Lemma C.1.7 we see that $K' \subset K$. On the other hand, again using Lemma C.1.7, we know that K is a compositum of extensions of the form $k(\sqrt[m]{\alpha})$. We may further assume that α is an algebraic integer and that $\mathrm{ord}_v(\alpha) \equiv 0 \pmod{m}$ for all places $v \notin S$, say $\mathrm{ord}_v(\alpha) = mr_v$ for some integer r_v.

For $v \notin S$, let $\mathfrak{p}_v$ be the prime ideal of $R_{k,S}$ corresponding to v, and let β be a generator of the (automatically principal) ideal

$$\prod_{v \notin S} \mathfrak{p}_v^{r_v}.$$

It follows that the element $\alpha' := \alpha\beta^{-m}$ is an S-integer and that $\mathrm{ord}_v(\alpha') = 0$ for all $v \notin S$. In other words, $\alpha' \in R^*_{k,S}$, and hence $k(\sqrt[m]{\alpha}) \subset K'$. Since the compositum of these fields is equal to K, we obtain the other inclusion $K \subset K'$, and so $K = K'$.

The above description of $K = K'$ makes it clear that K is generated by taking the mth roots of a set of generators of the group

$$R^*_{k,S}/(R^*_{k,S})^m,$$

and elementary Kummer theory tells us that $\mathrm{Gal}(K/k)$ is isomorphic to this quotient group. Finally, Dirichlet's unit theorem (see Theorem C.3.3 below) says that $R^*_{k,S}$ is the product of a finite cyclic group and a free group of rank $r(S)$, so our assumption that k contains an mth root of unity implies that the quotient $R^*_{k,S}/(R^*_{k,S})^m$ is isomorphic to $(\mathbb{Z}/m\mathbb{Z})^{r(S)+1}$. This completes the description of $\mathrm{Gal}(K/k)$. We also note that the fact that $R_{k,S}$ becomes principal after enlarging S is an easy corollary of Theorem C.3.1 below. □

We are now ready to prove our main result.

Theorem C.1.9. *Let A be an abelian variety of dimension g defined over a number field k, and fix an integer $m \geq 2$. Suppose that the m-torsion points of A are k-rational. Let S be a finite set of finite places of k that contains all places dividing m and all places of bad reduction of A. Assume further that the ring of S integers $R_{k,S}$ is principal. Then*

$$\operatorname{rank} A(k) \leq 2g \operatorname{rank} R^*_{k,S} = 2g(r_1 + r_2 + |S| - 1),$$

where r_1 and r_2 are as in (C.1.8).

PROOF. Let $r(S) = \operatorname{rank}(R^*_{k,S})$. By combining the results of Proposition C.1.2(ii), Proposition C.1.5, and Corollary C.1.8, we get an injection

$$\begin{aligned} A(k)/mA(k) &\hookrightarrow \operatorname{Hom}(\operatorname{Gal}(L/k), A_m) \\ &\cong \operatorname{Hom}\left((\mathbb{Z}/m\mathbb{Z})^{1+r(S)}, (\mathbb{Z}/m\mathbb{Z})^{2g}\right). \end{aligned}$$

We are going to compare the number of elements in these finite groups. Clearly,

$$\#\operatorname{Hom}\left((\mathbb{Z}/m\mathbb{Z})^{1+r(S)}, (\mathbb{Z}/m\mathbb{Z})^{2g}\right) = m^{2g(1+r(S))}.$$

On the other hand, since $A_m \subset A(k)$ by assumption, and since $\#A_m = m^{2g}$, we have

$$\#A(k)/mA(k) = m^{2g+\operatorname{rank} A(k)}.$$

This immediately gives the desired upper bound for $\operatorname{rank} A(k)$. □

C.2. The Kernel of Reduction Modulo p

In this section we give a proof of Theorem C.1.4 using the theory of formal groups. Other methods of proof are described in the exercises. Thus Exercise C.9 uses Hensel's lemma to show that, at least if we assume that our ring R is complete, the reduction map $A_m \to \tilde{A}_m$ is onto, and hence that it is injective, because both finite groups have the same cardinality. Exercise C.10 describes a scheme-theoretic proof, which is perhaps the most natural, but it demands considerably more in the way of prerequisites.

Our strategy in this section is to develop the rudiments of the theory of formal groups and to show that the kernel of reduction may be identified with the points of some formal group. Since it is easy to check that a (commutative) formal group has no prime-to-p torsion, this immediately gives the desired injectivity.

It is illuminating to observe that the analogous statement for the multiplicative group is both true and easy to verify. Thus let ξ and ξ' be distinct roots of unity in R^* of order prime to p, and let $\mathfrak{M}$ be a maximal

ideal of R with residue characteristic p. Then ξ and ξ' remain distinct when reduced modulo $\mathfrak{M}$. A quick proof uses the fact that if η is a root of unity of order n, then $1-\eta$ is a unit except when $n = p^m$, and even in the latter case it is still a p-unit. Another proof proceeds by directly showing that the only torsion in the kernel of reduction $1+\mathfrak{M}$ is p-power torsion. (See Proposition C.2.5 below). Similar elementary statements can be proven for the group $\mathrm{GL}(n)$ (Exercise C.14).

The reader will notice a close analogy between the exact sequence

$$0 \to 1+\mathfrak{M} \to \mathbb{G}_m(R) \to \mathbb{G}_m(\tilde{k})$$

and the exact sequence

$$0 \to A_1(k) \to A(k) \to \tilde{A}(\tilde{k})$$

that defines $A_1(k) = A_1(R)$.

We start now the construction of the formal group associated to an abelian variety A (or in fact any algebraic group) defined over a field k. Further motivation for this definition will be given later.

Let e be the identity element of the group A, and let $\widehat{\mathcal{O}}_{e,A}$ denote the completion of the local ring $\mathcal{O}_{e,A}$ of A at e with respect to its maximal ideal $\mathcal{M}_{e,A}$. We fix local parameters $x_1, \dots, x_g$ on A at e, which gives an isomorphism

$$\widehat{\mathcal{O}}_{e,A} \cong k[\![x_1, \dots, x_g]\!]$$

of the completed local ring with the ring of formal power series in g variables. (See Exercise A.1.12 or Shafarevich [1,II.2.2, Theorem 5].) The isomorphism is induced by the injection

$$\mathcal{O}_{e,A} \hookrightarrow k[\![x_1, \dots, x_g]\!],$$

which associates to each function its Taylor expansion at e with respect to the parameters $x_1, \dots, x_g$.

Next we consider the product $A \times A$, and for local parameters at the point $(e,e) \in A \times A$ we choose the functions $y_1, \dots, y_g, z_1, \dots, z_g$, where

$$y_i := x_i \circ p_1 \qquad \text{and} \qquad z_i := x_i \circ p_2.$$

Just as above, this choice furnishes us with an isomorphism

$$\widehat{\mathcal{O}}_{(e,e),A\times A} \cong k[\![y_1, \dots, y_g, z_1, \dots, z_g]\!]$$

for the completed local ring of $A \times A$ at the point (e,e).

Now consider the addition map

$$\mathrm{add} : A \times A \to A$$

giving the group law on A. It induces a map of local rings

$$\text{add}^* : \mathcal{O}_{e,A} \longrightarrow \mathcal{O}_{(e,e),A\times A},$$

and hence via the above isomorphisms, a map

$$\text{add}^* : k[\![x_1,\dots,x_g]\!] \longrightarrow k[\![y_1,\dots,y_g,z_1,\dots,z_g]\!]$$

of formal power series rings. To ease notation, we let

$$F_i := \text{add}^*(x_i),$$

so the F_i's should "look like" the coordinates of the group law on A. The following lemma makes this precise. (Note that in the interest of conserving variable names, we have relabeled the variables in the F_i's.)

Lemma C.2.1. *Let $F_1,\dots,F_g \in [\![X_1,\dots,X_g,Y_1,\dots,Y_g]\!]$ be the formal power series associated to the group law on an abelian variety for a fixed choice of local parameters at the origin as described above. Then the g-tuple of formal power series $F=(F_1,\dots,F_g)$ defines a commutative formal group (of dimension g). That is, it satisfies the following conditions:*

(1) $F(X,Y) = X + Y + (\textit{terms of degree} \geq 2)$ [infinitesimal group]

(2) $F(X,F(Y,Z)) = F(F(X,Y),Z)$ [associativity]

(3) $F(X,Y) = F(Y,X)$ [commutativity]

(4) $F(X,0) = X$ *and* $F(0,Y) = Y$ [neutral element]

(5) *There exists a unique g-tuple of formal series without constant term $i(X) = (i_1(X),\dots,i_g(X))$ such that $F(X,i(X)) = F(i(X),X) = 0$.* [inverse]

PROOF. Intuitively, all these properties are infinitesimal translations of the properties of the addition law on A. For example, the differential of the addition map is given by $(X,Y)\mapsto X+Y$, from which property (1) follows. To obtain properties (2) and (3), we just use the formulas

$$\text{add}(\text{add}(x,y),z) = \text{add}(x,\text{add}(y,z)) \qquad \text{and} \qquad \text{add}(x,y) = \text{add}(y,x),$$

which say that A is an abelian group. Finally, it is relatively straightforward to deduce properties (4) and (5) from properties (1), (2), and (3). □

It is a rather simple matter to compute the formal groups of the additive group $\mathbb{G}_a$ and the multiplicative group $\mathbb{G}_m$, and even of the general linear group $\text{GL}(n)$. Thus

$$F_{\mathbb{G}_a}(X,Y) = X+Y \qquad \text{and} \qquad F_{\mathbb{G}_m}(X,Y) = X+Y+XY.$$

The formal group of $\text{GL}(n)$, which is not commutative when $n \geq 2$, is given by the coordinate functions

$$F_{ij}(X,Y) := X_{ij} + Y_{ij} + \sum_{h=1}^{n} X_{ih}Y_{hj}.$$

By way of contrast, there is no such simple formula for the formal group of an abelian variety, not even in dimension one (i.e., for an elliptic curve).

Definition. A formal group is *defined over the ring* R if the coefficients of its defining power series F all lie in the ring R.

Definition. Let $F = (F_1, \dots, F_g)$ and $G = (G_1, \dots, G_h)$ be formal groups defined over R. A *homomorphism from* F *to* G *defined over* R is an h-tuple of formal series without constant terms $f = (f_1, \dots, f_h) \in R[\![X_1, \dots, X_g]\!]$ with the property that

$$G(f(X), f(Y)) = f(F(X,Y)).$$

The homomorphism f is an *isomorphism over* R if there exists a g-tuple of formal series without constant terms $f' = (f'_1, \dots, f'_g) \in R[\![X_1, \dots, X_h]\!]$ such that $f(f'(X)) = f'(f(X)) = X$.

Notice that if F and G are two formal groups defined over a ring R, and if R is a subring of a larger ring k (e.g., k could be a field), then F and G may be isomorphic over k, but nonisomorphic over R. The next lemma describes a simple, but extremely important, way to determine whether a map between formal groups is an isomorphism.

Lemma C.2.2. *Let* $f = (f_1, \dots, f_g) \in R[\![X_1, \dots, X_g]\!]$ *be a* g*-tuple of formal series without constant terms, say*

$$f_i = \sum_{j=1}^{g} f_{ij} X_j + (\text{terms of degree } \geq 2).$$

Form the matrix (f_{ij}) *whose entries are the coefficients of the linear terms of the* f_i*'s. If* $\det(f_{ij})$ *is a unit in* R*, then there exists a* g*-tuple of power series* $f' = (f'_1, \dots, f'_g) \in R[\![X_1, \dots, X_g]\!]$ *without constant terms such that*

$$f(f'(X)) = f'(f(X)) = X.$$

Conversely, if $\det(f_{ij})$ *is not a unit in* R*, then no such power series exists. The inverse power series* f' *is often denoted by* f^{-1}*.*

PROOF. Easy (see Exercise C.11). □

Let $f = (f_1, \dots, f_g)$ be as in the previous lemma and let F be a formal group over R. Then it is trivial to verify that $G := f^{-1}(F(f(X), f(Y)))$ defines a formal group isomorphic over R to F. If A is an abelian variety defined over a field k, then the various formal groups (depending on the choice of local parameters) associated to A are all isomorphic over k, since the determinant of the change-of-variable matrix is a unit in k.

A natural and important example of a homomorphism of commutative formal groups is provided by "multiplication-by-m," which is described

inductively for all integers m by the following formulas. (Note that $i(X)$ is the inverse power series described in Lemma C.2.1.)

$$[-1](X) = i(X), \qquad [0](X) = 0, \qquad [1](X) = X,$$
$$[m](X) = F\big(X, [m-1](X)\big), \qquad [m](X) = F\big(i(X), [m+1](X)\big).$$

Lemma C.2.3. *Let F be a commutative formal group defined over R, let m be an integer, and let $[m]$ be the array of formal power series as defined above.*
(a) The array of power series $[m]$ gives an endomorphism of the formal group F defined over R.
(b) The endomorphism $[m]$ is an isomorphism over R if and only if m is a unit in R.

PROOF. The fact that $[m]$ is an endomorphism follows from the associativity of F and induction on m. It is also immediate by induction that $[m](X) = mX + \cdots$, so the matrix of linear coefficients for $[m]$ is m times the identity matrix. Since $\det(mI) = m^g$, it follows from Lemma C.2.2 that $[m]$ has an inverse over R if and only if m is a unit in R. □

The previous considerations were purely geometric. We now introduce arithmetic by taking R to be a complete local valuation ring with maximal ideal $\mathcal{M}$, fraction field $k = \operatorname{Frac}(R)$, and residue field $\tilde{k} = R/\mathcal{M}$.

Our first important observation is that if an abelian variety has good reduction, then it is possible to select local parameters so that the associated formal group has coefficients in R. This proof uses the property that for an abelian variety A defined over a number field, good reduction at some prime includes the condition that addition also has good reduction, i.e., that the addition map on A reduces to a morphism

$$\widetilde{\text{add}} : \tilde{A} \times \tilde{A} \longrightarrow \tilde{A}.$$

Lemma C.2.4. *Let R be a local ring as above, let A/k be an abelian variety with good reduction at $\mathcal{M}$, and let $\tilde{A}/\tilde{k}$ denote the reduced abelian variety. Let $x_1, \dots, x_g$ be local parameters on A at e with the property that their reductions $\tilde{x}_1, \dots, \tilde{x}_g$ are local parameters on $\tilde{A}$ at $\tilde{e}$. Let F be the formal group of A with respect to the parameters $x_1, \dots, x_g$. Then $F_i \in R[\![X_1, \dots, X_g, Y_1, \dots, Y_g]\!]$; that is, the coefficients of the formal group power series lie in the ring R.*

PROOF. Applying the above construction to the local parameters $\tilde{x}_1, \dots, \tilde{x}_g$ will yield power series $G_i \in \tilde{k}[\![X_1, \dots, X_g, Y_1, \dots, Y_g]\!]$ giving the group law on $\tilde{A}$, and these power series must equal the reduction modulo $\mathcal{M}$ of the power series F_i. Hence the F_i's must have integral coefficients, i.e., coefficients in R.

□

In general, a formal group does not define an actual group. A useful intuition is that a formal group is a group law without any group elements! However, although a formal group is merely a template for a group, if we substitute values for its variables and if we can attach a reasonable meaning to the resulting infinite series, then all of the group axioms will automatically be true.

For example, suppose that R is a complete local ring with maximal ideal $\mathcal{M}$ as above, and suppose that F is a formal group over R. Then the series $F_i(X,Y)$ will converge in R for any choice of $X,Y \in \mathcal{M}^g$. (Note that this means that X and Y are g-tuples of elements of $\mathcal{M}$, not the gth power of elements of $\mathcal{M}$.) In this way the formal group F defines a group structure on $\mathcal{M}^g$.

Definition. Let F be a formal group of dimension g defined over a complete local ring R with maximal ideal $\mathcal{M}$. The *group associated to F/R*, denoted by $F(\mathcal{M})$, is the set of g-tuples $\mathcal{M}^g$ with the group law

$$\mathcal{M}^g \times \mathcal{M}^g \xrightarrow{+_F} \mathcal{M}^g, \qquad X +_F Y := F(X,Y).$$

Example C.2.4.1. The formal group associated to $\mathbb{G}_a$ is simply $\mathcal{M}$ with its usual addition. The formal group associated to $\mathbb{G}_m$ is isomorphic to the set $1+\mathcal{M}$ with its usual multiplication.

Proposition C.2.5. *Let R be a complete local ring with maximal ideal $\mathcal{M}$ and residue characteristic p, and let F be a formal group over R. Then the group $F(\mathcal{M})$ has no prime-to-p torsion.*

PROOF. Let m be an integer not divisible by p. Then m is a unit in R, so Lemma C.2.3 tells us that the series $[m](X) = mX + \cdots$ has a formal inverse $[m]^{-1} = m^{-1}X + \cdots$ with coefficients in R. Both series are convergent when applied to elements in $\mathcal{M}^g$, and therefore multiplication by m is an automorphism of the group $F(\mathcal{M})$ (i.e., an isomorphism from $F(\mathcal{M})$ to itself). In particular, multiplication-by-m has trivial kernel, so $F(\mathcal{M})$ has no m-torsion. □

Theorem C.2.6. *Let R be a complete local ring with maximal ideal $\mathcal{M}$, fraction field k, and residue field $\tilde{k}$. Let A/k be an abelian variety having good reduction at $\mathcal{M}$, and let*

$$A_1(k) := \ker\left\{A(k) \xrightarrow{\text{red}} \tilde{A}(\tilde{k})\right\}$$

be the kernel of reduction. Further, let F be the formal group of A as described in Lemma C.2.4, so in particular, F is defined over R. Then there is an isomorphism

$$F(\mathcal{M}) \cong A_1(k).$$

PROOF. We show that there are injective maps

$$F(\mathcal{M}) \longrightarrow A_1(k) \qquad \text{and} \qquad A_1(k) \longrightarrow F(\mathcal{M})$$

such that their composition is the identity, and such that the first map is a homomorphism. These two facts then imply that both maps are isomorphisms.

Choose local coordinates $x_1, \ldots, x_g$ for A at e as in Lemma C.2.4, and let U be an open affine neighborhood of e on which the x_i's give coordinates. This means we can write the affine coordinate ring of U as $k[U] = k[x_1, \ldots, x_n]$ with $x_{g+1}, \ldots, x_n$ expressible as power series

$$x_j = f_j(x_1, \ldots, x_g) \in R[\![x_1, \ldots, x_g]\!].$$

We can make these choices so that the same formulas hold when we reduce modulo $\mathcal{M}$, that is,

$$\tilde{x}_j = \tilde{f}_j(\tilde{x}_1, \ldots, \tilde{x}_g).$$

In particular, the f_j's have coefficients in R.

Now the map given by the formula

$$(X_1, \ldots, X_g) \longmapsto (X_1, \ldots, X_g, f_{g+1}(X_1, \ldots, X_g), \ldots, f_n(X_1, \ldots, X_g))$$

provides the first map $F(\mathcal{M}) \to A_1(k)$, and it is clear that this map is injective. Further, it is clear from the definition of the formal group that this map is a homomorphism.

Next consider the projection $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_g)$ giving the second map $A_1(k) \to F(\mathcal{M})$. The fact that $x_1, \ldots, x_g$ are local parameters on U imply that this map is injective, which completes the proof of Theorem C.2.6. □

We conclude this section by observing that Proposition C.2.5 and Theorem C.2.6 provide us with the promised proof of Theorem C.1.4.

C.3. Appendix: Finiteness Theorems in Algebraic Number Theory

We give a short introduction to the geometry of numbers in order to prove the following three basic finiteness results from algebraic number theory.

Theorem C.3.1. *The group of ideal classes of a number field is finite.*

Theorem C.3.2. (Hermite) *The set of number fields (viewed as subfields of $\mathbb{C}$) with given discriminant is finite.*

Theorem C.3.3. (Dirichlet unit theorem) *Let k be a number field, let r_1 and r_2 be respectively the number of its real and complex archimedean*

places of k, and let S be a finite set of nonarchimedean places of k. Then the group of S-units

$$R_{k,S}^* := \{x \in k^* \mid |x|_v = 1 \text{ for all finite } v \notin S\}$$

is a finitely generated group of rank

$$\operatorname{rank} R_{k,S}^* = r(S) = r_1 + r_2 - 1 + \#S.$$

Let us set some notation for the rest of this section.

n or n_k	the degree $[k:\mathbb{Q}]$ of a number field k.
r_1, r_2	the number of real, resp. complex, places of k.
$\sigma_1, \ldots, \sigma_{r_1}$	the distinct real embeddings $k \hookrightarrow \mathbb{R}$.
$\sigma_{r_1+1}, \ldots, \sigma_{r_1+r_2}$, $\bar{\sigma}_{r_1+1}, \ldots, \bar{\sigma}_{r_1+r_2}$	the distinct complex embeddings $k \hookrightarrow \mathbb{C}$.
Δ or Δ_k	the absolute value of the discriminant of $k/\mathbb{Q}$.

We will also use the fact that, if $\alpha_1, \ldots, \alpha_n$ is a $\mathbb{Z}$-basis of R_k, then

$$\Delta = \left|\det\big(\sigma_i(\alpha_j)\big)^2\right|.$$

We begin by introducing the *canonical embedding* of k,

$$\sigma : k \longrightarrow E_\infty := \prod_{v \in M_k^\infty} k_v \cong \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}, \qquad \sigma(x) = \big(\sigma_1(x), \ldots, \sigma_{r_1+r_2}(x)\big).$$

We remind the reader that a discrete subgroup Γ of a real vector space E of dimension n is isomorphic to $\mathbb{Z}^r$ for some integers $r \le n$, and that Γ is said to be a *lattice* if $r = n$. The *volume of a lattice* Γ is defined to be the volume of any fundamental domain for Γ. If $e_1, \ldots, e_n$ is a $\mathbb{Z}$-basis for Γ, then the volume of Γ is given by $\left|\det(e_1, \ldots, e_n)\right|$.

Lemma C.3.4. *The image of the ring of integers $\sigma(R_k)$ inside E_∞ is a lattice of volume $2^{-r_2}\sqrt{\Delta_k}$. The volume of the image $\sigma(I)$ of a nonzero ideal I of R_k is $2^{-r_2}\sqrt{\Delta_k}N(I)$.*

PROOF. Let $\alpha_1, \ldots, \alpha_n$ be a basis of R_k. Making a change of variables from $z, \bar{z}$ to $\operatorname{Re}(z), \operatorname{Im}(z)$ introduces a factor 2, so remembering that we have

$$\sigma_{r_1+j}(\alpha_i) = \operatorname{Re}\sigma_{r_1+j}(\alpha_i) + \sqrt{-1}\operatorname{Im}\sigma_{r_1+j}(\alpha_i),$$

we see that the volume of the cube generated by the $\sigma(\alpha_i)$ is

$$\left|\big(2\sqrt{-1}\big)^{-r_2}\det\big(\sigma_i(\alpha_j)\big)\right| = 2^{-r_2}\sqrt{\Delta_k}.$$

The second claim is then clear, since $\sigma(I)$ has index $N(I)$ inside $\sigma(R_k)$. □

Corollary C.3.5. *(Minkowski) There exists a constant $c_1 > 0$ such that every nonzero ideal I of R_k contains a nonzero element $\alpha \in I$ with norm satisfying*

$$\left|N_{\mathbb{Q}}^k(\alpha)\right| \le c_1\sqrt{\Delta_k}\cdot N(I).$$

Further, the constant c_1 can be chosen to be $c_1 = (4/\pi)^{r_2}n!/n^n$.

PROOF. For any real number $t \ge 0$, consider the symmetric convex set

$$\mathcal{C}_t = \left\{(y,z)\in\mathbb{R}^{r_1}\times\mathbb{C}^{r_2}\;\middle|\;\sum_{i=1}^{r_1}|y_i| + 2\sum_{i=r_1+1}^{r_1+r_2}|z_j| \le t\right\}.$$

By homogeneity, it is clear that $\mathcal{C}_t$ has volume c_0t^n for some constant c_0 independent of t, and a straightforward computation shows that $c_0 = 2^{r_1}(\pi/2)^{r_2}/n!$. We choose t to satisfy

$$c_0t^n = 2^n\cdot\text{Volume}\big(\sigma(I)\big).$$

Then Minkowski's theorem (B.5.4) tells us that there exists a nonzero element $\alpha \in I$ satisfying $\sigma(\alpha)\in\mathcal{C}_t$. Using the arithmetic–geometric inequality, we can bound the norm of α as follows:

$$\begin{aligned}
\left|N_{\mathbb{Q}}^k(\alpha)\right| &= \prod_{i=1}^{r_1}\left|\sigma_i(\alpha)\right| \times \prod_{j=1}^{r_2}\left|\sigma_{r_1+j}(\alpha)\right|^2\\
&\le \left(\frac{1}{n}\Big(\sum_{i=1}^{r_1}|\sigma_i(\alpha)| + 2\sum_{j=1}^{r_2}|\sigma_{r_1+j}(\alpha)|\Big)\right)^n\\
&\le \frac{t^n}{n^n} \qquad \text{since } \sigma(\alpha)\in\mathcal{C}_t\\
&= \frac{2^n}{c_0\cdot n^n}\cdot\text{Volume}\big(\sigma(I)\big) \qquad \text{from our choice of } t.
\end{aligned}$$

Now plugging in the value of c_0 from above and the volume of $\sigma(I)$ from Lemma C.3.4 gives the desired result. □

Lemma C.3.6. *The set of ideals of a given norm is finite.*

PROOF. Let $m \ge 1$ be an integer. The quotient ring R_k/mR_k is finite (indeed, as an additive group it is isomorphic to $(\mathbb{Z}/m\mathbb{Z})^n$), so R_k/mR_k contains only a finite number of ideals. The ideals in R_k containing the ideal mR_k correspond to the ideals of R_k/mR_k, so there are only finitely many ideals in R_k containing mR_k. Finally, we observe that if $N(I) = m$,

then $mR_k \subset I$. Therefore, there are only finitely many ideals in R_k of a given norm. □

Lemma C.3.7. *Every ideal class of k contains an ideal of norm less than $c_1\sqrt{\Delta_k}$, where c_1 is as in (C.3.5).*

PROOF. Let C be an ideal class, and let I' be an ideal of R_k in the inverse ideal class $C' = C^{-1}$. Corollary C.3.5 tells us that there exists a nonzero element $\alpha' \in I'$ satisfying $|N_{\mathbb{Q}}^k(\alpha')| \le c_1\sqrt{\Delta_k}N(I')$. Then $I := \alpha' I'^{-1}$ is an ideal in the ideal class C, and we can bound its norm by

$$N(I) = |N_{\mathbb{Q}}^k(\alpha')| \cdot N(I')^{-1} \le c_1\sqrt{\Delta}.$$

This shows that every ideal class contains an ideal of norm at most $c_1\sqrt{\Delta}$. □

Notice that Theorem C.3.1 is now an immediate consequence of Lemmas C.3.6 and C.3.7, since the lemmas imply that the finitely many ideals of norm at most $c_1\sqrt{\Delta}$ represent all of the ideal classes, and hence the ideal class group is finite. We also remark that if the ideals $I_1, \dots, I_h$ represent the distinct ideal classes, and if S is the set of prime ideals dividing $I_1 \cdots I_h$, then the ring $R_{k,S}$ will be principal.

Lemma C.3.8. *There exists an absolute constant c_2 such that every number field $k/\mathbb{Q}$ satisfies*

$$[k : \mathbb{Q}] \le c_2 \log \Delta_k.$$

PROOF. From the previous proof we extract the estimate

$$1 \le \left(\frac{4}{\pi}\right)^{r_2} \frac{n!}{n^n}\sqrt{\Delta}.$$

Using $r_2 \le n/2$ and $n^n/n! \ge (9/4)^{n-1}$, we obtain after some calculation that $\Delta \ge \frac{4}{9}(3\pi/4)^n$. Taking logarithms gives the desired result. □

PROOF (of Hermite's theorem (C.3.2)). Let k be a number field with given discriminant $\Delta_k = \Delta$. Using the previous lemma (C.3.8), we may assume that the degree n, and even its type (r_1, r_2), is fixed. Minkowski's theorem (B.5.4) says that there is a T, depending only on r_1, r_2, and Δ, such that there exists an element $\alpha \in R_k$ satisfying:

(i) $|\sigma_i(\alpha)| \le \frac{1}{2}$ for all $2 \le i \le r_1 + r_2$;

(ii) $\begin{cases} |\sigma_1(\alpha)| \le T & \text{if } r_1 > 0, \\ |\operatorname{Re}(\sigma_1(\alpha))| \le \frac{1}{2} \quad \text{and} \quad |\operatorname{Im}(\sigma_1(\alpha))| \le T & \text{if } r_1 = 0. \end{cases}$

We now show that $k = \mathbb{Q}(\alpha)$. First we note that

$$1 \le \left|N_{\mathbb{Q}}^{k}(\alpha)\right| \le \prod_{v \in M_k^\infty} |\alpha|_v \le |\sigma_1(\alpha)|.$$

It follows that $\sigma_1(\alpha)$ is different from all of the other conjugates $\sigma_i(\alpha)$ except (perhaps) from $\bar{\sigma}_1(\alpha)$ if $r_1 = 0$ and $\sigma_1(\alpha)$ is real. But in the latter case,

$$|\operatorname{Im}(\sigma_1(\alpha))| \ge |\sigma_1(\alpha)| - |\operatorname{Re}(\sigma_1(\alpha))| \ge \frac{1}{2},$$

so $\sigma_1(\alpha)$ cannot be real. Hence $\sigma_1(\alpha) \ne \sigma_i(\alpha)$ for all $i \ge 2$.

Now suppose that $\sigma_i(\alpha) = \sigma_j(\alpha)$ for some i, j. Then the map $\sigma_k = \sigma_1\sigma_i^{-1}\sigma_j$ satisfies $\sigma_1(\alpha) = \sigma_k(\alpha)$, so we have $\sigma_k = \sigma_1$, and hence $\sigma_i = \sigma_j$. This proves that the images $\sigma_i(\alpha)$ for $1 \le i \le n$ are distinct, which implies that $[\mathbb{Q}(\alpha) : \mathbb{Q}] \ge n$, and hence that $k = \mathbb{Q}(\alpha)$.

Now consider the coefficients of the minimal polynomial of α over $\mathbb{Z}$. These coefficients are symmetric functions in the $\sigma_i(\alpha)$'s, and thus their size is bounded by a function of r_1, r_2, and Δ. (Remember that T is chosen in terms of these three quantities.) Thus there are only finitely many possibilities for the coefficients of the minimal polynomial of α, which completes the proof that there are only finitely many fields with a given discriminant. □

We now introduce another lattice associated to a number field. Let S be a finite set of nonarchimedean places of k with $s = \#S$. We set

$$T = S \cup M_k^\infty \qquad \text{and} \qquad t = \#T = r_1 + r_2 + s.$$

We define a map

$$k^* \longrightarrow \mathbb{R}^t, \qquad x \longmapsto \left(\log \|x\|_v\right)_{v \in T},$$

and we denote by Φ the restriction of this map to $R_{k,S}^*$. (Recall that the group of S-units $R_{k,S}^*$ is the group of elements of k^* such that $|\alpha|_v = 1$ for all $v \notin T$.) The map Φ is often called the *regulator map* (or the *S-regulator map* if S is not empty). It is also sometimes called the *logarithmic embedding*, although as the next result shows, Φ actually gives only an embedding of $R_{k,S}^*/\mu_k$ (i.e., modulo torsion).

Lemma C.3.9. *Let $\Phi : R_{k,S}^* \to \mathbb{R}^t$ be the S-regulator map described above.*
(a) *The kernel of Φ is μ_k, the group of roots of unity in k^*.*
(b) *The image of Φ is a discrete subgroup contained in the hyperplane $\sum_{v \in T} x_v = 0$.*

PROOF. An element α of $R_{k,S}^*$ is in the kernel of Φ if and only if $|\alpha|_v = 1$ for all places $v \in M_k$. By Kronecker's theorem (B.2.3.1), this is equivalent to saying that α is a root of unity, which proves (a).

Next we observe that if $\Phi(\alpha)$ is confined to a bounded set, then the height of α is bounded, and so Theorem B.2.3 tells us that α may take on only finitely many values. This shows that the image $\Phi(R_{k,S}^*)$ is discrete. Finally, the product formula implies that every $\alpha \in R_{k,S}^*$, satisfies $\sum_{v\in T} \log \|\alpha\|_v = 0$, which proves that $\Phi(R_{k,S}^*)$ lies in the hyperplane $\sum_{v\in T} x_v = 0$. □

Notice that Lemma C.3.9 and the basic fact that a discrete subgroup of $\mathbb{R}^n$ is free of rank at most n already tells us that

$$R_{k,S}^* = \mu_k \times \mathbb{Z}^{r(S)} \qquad \text{with} \qquad r(S) \le r_1 + r_2 - 1 + s.$$

This inequality is actually all that we need for our applications, but for the sake of completeness we will prove that $r(S) = r_1+r_2-1+s$ in the case that $S = \emptyset$. We leave the general case, which is proven similarly, as an exercise for the reader. We also note that by definition, the *regulator* of the number field k is the volume of the lattice $\Phi(R_k^*)$ in the $(r_1 + r_2 - 1)$-dimensional hyperplane in which it lies.

PROOF (of Dirichlet's unit theorem (C.3.3) when $S = \emptyset$) Choose any element

$$u = (u_v)_{v\in M_k^\infty} \in E_\infty \quad \text{satisfying} \quad N(u) := \prod_{v\in M_k^\infty} |u_v|^{n_v} = 1.$$

Then multiplication by u is a linear transformation of E_∞ with determinant of absolute value one. It follows that the lattice $u\sigma(R_k)$ has the same volume as the lattice $\sigma(R_k)$, namely $2^{-r_2}\sqrt{\Delta}$ (C.3.4). For any constants $t_1, \dots, t_{r_1+r_2}$ such that $T := t_1 \cdots t_{r_1} t_{r_1+1}^2 \cdots t_{r_1+r_2}^2$ is large enough, Minkowski's theorem (B.5.4) gives a nonzero element $\alpha \in R_k$ such that the element $x = u\sigma(\alpha)$ satisfies $|x_v| \le t_v$ for all $v \in M_k^\infty$. We deduce that $|N_{\mathbb{Q}}^k(\alpha)| = N(x) \le T$. By Lemma C.3.6, there exists a finite set $\alpha_1, \dots, \alpha_N \in R_k$ such that all $\alpha \in R_k$ with $|N_{\mathbb{Q}}^k(\alpha)| \le T$ may be written $\alpha = \alpha_i\varepsilon$ for some $1 \le i \le N$ and $\varepsilon \in R_k^*$. We may therefore write $x = u\sigma(\alpha_i\varepsilon)$, or equivalently, $u = x\sigma(\alpha_i^{-1})\sigma(\varepsilon^{-1})$. Taking the logarithms of the absolute values, we obtain

$$\log |u_v| = \log |x_v| - \log |\alpha_i|_v - \log |\varepsilon|_v.$$

Note that every point in the hyperplane $\sum_{v\in T} x_v = 0$ in $\mathbb{R}^{r_1+r_2}$ can be represented by an element $(\log |u_v|)_{v\in M_k^\infty}$, while the elements

$$\left(\log |x_v| - \log |\alpha_i|_v\right)_{v\in M_k^\infty}$$

lie in a bounded set. (It is clear that the $|x_v|$'s are bounded above, and their product is bounded below, so they are also individually bounded below.) This proves that the hyperplane is the union of translates of a bounded set by elements of $\Phi(R_k^*)$. Therefore $\Phi(R_k^*)$ is a lattice (i.e., has maximal rank). □

C.4. Appendix: The Selmer and Tate–Shafarevich Groups

We return to the proof of the weak Mordell–Weil theorem and analyze it in a more abstract setting, using Galois cohomology. For the relevant definitions and properties of H^1, cocycles, and coboundaries, see Section C.5 below.

Let $G = G_k$ be the Galois group $\mathrm{Gal}(\bar{k}/k)$, and let A be an abelian variety defined over k. For each $x \in A(k)$, choose a point $y \in A(\bar{k})$ satisfying $my = x$ and define as before the map

$$t(\,\cdot\,, y) : G \longrightarrow A_m, \qquad \sigma \longmapsto y^\sigma - y.$$

We do not assume, as we did in Section C.1, that G acts trivially on A_m, so the formula of Proposition C.1.1 becomes

$$t(\sigma\sigma', y) = t(\sigma', y)^\sigma + t(\sigma, y).$$

In other words, the map $\sigma \mapsto t(\sigma, y)$ is a cocycle $G \to A_m$.

Now suppose that we choose some other y' such that $my' = x$. Then the point $b := y' - y$ is in A_m, and

$$t(\sigma, y') - t(\sigma, y) = (y'^\sigma - y') - (y^\sigma - y) = (y' - y)^\sigma - (y' - y) = b^\sigma - b.$$

Thus the difference $t(\,\cdot\,, y') - t(\,\cdot\,, y')$ is a coboundary, so the cohomology class of $t(\,\cdot\,, y)$ in $H^1(G, A_m)$ depends only on x, independent of the choice of y. In other words, we get a well-defined map $\delta : A(k) \to H^1(G, A_m)$. The next proposition gives a slight generalization of this construction.

Proposition C.4.1. *Let $\alpha : A \to B$ be an isogeny of two abelian varieties defined over k. Then the short exact sequence*

$$0 \longrightarrow \ker(\alpha) \xrightarrow{\ \iota\ } A(\bar{k}) \xrightarrow{\ \alpha\ } B(\bar{k}) \longrightarrow 0$$

induces a long exact sequence of cohomology groups

$$\begin{aligned} 0 \longrightarrow \ker(\alpha)(k) &\xrightarrow{\ \iota\ } A(k) \xrightarrow{\ \alpha\ } B(k) \\ &\xrightarrow{\ \delta\ } H^1(G, \ker(\alpha)) \xrightarrow{\ \iota\ } H^1(G, A(\bar{k})) \xrightarrow{\ \alpha\ } H^1(G, B(\bar{k})). \end{aligned}$$

The connecting homomorphism δ is defined as follows: Let $x \in B(k)$, and select $y \in A(\bar{k})$ such that $\alpha(y) = x$. Then define $\delta(x)$ to be the cohomology class associated to the cocycle

$$\delta(x) : G \longrightarrow \ker(\alpha), \qquad \delta(x)(\sigma) = y^\sigma - y.$$

The above long exact sequence gives rise to the following fundamental short exact sequence,

$$0 \longrightarrow B(k)/\alpha A(k) \xrightarrow{\ \delta\ } H^1(G, \ker(\alpha)) \longrightarrow H^1(G, A(\bar{k}))[\alpha] \longrightarrow 0,$$

where $H^1(G, A(\bar{k}))[\alpha]$ denotes the kernel of the map $\alpha : H^1(G, A(\bar{k})) \to H^1(G, B(\bar{k}))$.

PROOF. The existence of the long exact sequence comes from basic properties of group cohomology, and the definition of δ follows from the way that the long exact sequence in group cohomology is constructed from a short exact sequence of G-modules. The reader who is unfamiliar with this material should consult Section C.5 and, for further details, a book such as of Hilton–Stammbach [1]. □

It would be nice to use the injection $\delta : B(k)/\alpha A(k) \hookrightarrow H^1(G_k, \ker(\alpha))$ to prove directly the finiteness of $B(k)/\alpha A(k)$, but unfortunately the group $H^1(G_k, \ker(\alpha))$ may be infinite. For example, this is the case when k is a number field. Thus we must somehow cut down the size of $H^1(G_k, \ker(\alpha))$. One method to do this is via localization at primes.

Indeed, for each place v of k, let k_v be the completion of k at v and let $G_v := \mathrm{Gal}(\bar{k}_v/k_v)$ be the absolute Galois group of k_v. We may consider G_v to be a subgroup of G_k, and hence we obtain restriction maps $H^1(G_k, \cdot\,) \to H^1(G_v, \cdot\,)$. There is a local exact sequence analogous to the exact sequence in Proposition C.3.1, and the global exact sequence maps to the local exact sequence via restriction, yielding the following commutative diagram.

$$\begin{array}{ccccccccc}
0 & \to & B(k)/\alpha A(k) & \xrightarrow{\delta} & H^1(G_k, \ker(\alpha)) & \longrightarrow & H^1(G_k, A(\bar{k}))[\alpha] & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & B(k_v)/\alpha A(k_v) & \xrightarrow{\delta_v} & H^1(G_v, \ker(\alpha)) & \longrightarrow & H^1(G_v, A(\bar{k}_v))[\alpha] & \to & 0
\end{array}$$

Now observe that if $x \in B(k)$, then the restriction $\delta_v(x)$ will be in the kernel of the right-hand map. Turning this observation around, we see that each place v gives us some information about the image of δ in $H^1(G_k, \ker(\alpha))$. This remark motivates the definition of the following two groups. We will see later that these groups also admit a very interesting arithmetic/geometric interpretation.

Definition. Let $\alpha : A \to B$ be an isogeny of abelian varieties defined over a number field k. The *Selmer group of A with respect to α* is the group

$$\mathrm{Se}\ell^{(\alpha)}(A/k) := \bigcap_v \ker\{H^1(G_k, \ker(\alpha)) \to H^1(G_v, A(\bar{k}_v))[\alpha]\}.$$

The *Tate–Shafarevich group of A* is the group

$$\text{Ш}(A/k) := \bigcap_v \ker\{H^1(G_k, A(\bar{k})) \to H^1(G_v, A(\bar{k}_v))\}.$$

In both formulas, the product is taken over all places v of k.

From the exact sequence of Proposition C.4.1 and these definitions, we readily deduce the following important fundamental sequence:

$$0 \longrightarrow B(k)/\alpha A(k) \longrightarrow \mathrm{Se}\ell^{(\alpha)}(A/k) \longrightarrow Ш(A/k)[\alpha] \longrightarrow 0.$$

We claim that the Selmer group $\mathrm{Se}\ell^{(\alpha)}(A/k)$ is finite, which will certainly also imply the finiteness of $B(k)/\alpha A(k)$ and $Ш(A/k)[\alpha]$. We will prove the finiteness of $\mathrm{Se}\ell^{(\alpha)}(A/k)$ using a ramification argument very similar to the one that we already used to prove the weak Mordell–Weil theorem.

Before proving the finiteness of the Selmer group, we give a geometric interpretation of the groups $\mathrm{Se}\ell^{(\alpha)}(A/k)$ and $Ш(A/k)$. An element of $H^1(G_k, A(\bar{k}))$ corresponds to a principal homogeneous space. (See the next section for the definition, and especially Proposition C.5.3 for the correspondence between cohomology classes and homogeneous spaces.) The cohomology class is trivial if and only if the corresponding homogeneous space has a rational point. Thus we see that the homogeneous spaces corresponding to elements of the Selmer group $\mathrm{Se}\ell^{(\alpha)}(A/k)$ possess k_v-points for every place v of k. Similarly, nontrivial elements of the Tate–Shafarevich group $Ш(A/k)$ correspond to homogeneous spaces that have k_v rational points for every place v, yet nevertheless have no k-rational points. It is far from obvious that there can exist any such spaces for which the Hasse principle fails. An example is given in Exercise C.15. (See Silverman [1, Chapter X, Proposition 6.5] for some additional examples.) The existence of nontrivial elements of $Ш(A/k)$ accounts for the ineffectivity of the Mordell–Weil theorem.

As we will now see, not only is the Selmer group $\mathrm{Se}\ell^{(\alpha)}(A/k)$ finite, but it is also effectively computable (at least in principle). This is true because the question of whether a given variety has rational points for every completion k_v is computable in finite time by combining Hensel's lemma (see Exercise C.9) with estimates for the number of points on varieties over finite fields.

Thus the above exact sequence provides us with a finite collection of principal homogeneous spaces, each of which has k_v-rational points for all places v, and the only task remaining is to determine which of these homogeneous spaces have at least one k-rational point. Unfortunately, no algorithm is known that is guaranteed to determine whether or not a specific homogeneous space possesses a k-rational point. This is true even for curves of genus 1. (See, for example, the discussion in Tate [2].)

Showing that the Selmer group is finite amounts to proving the weak Mordell–Weil theorem and is done by using essentially the same ramification argument. First we must define what it means for a cohomology class to be unramified.

Definition. Let v be a place of k, and let $I_v \subset G_k$ be an inertia for v. A cohomology class $\phi \in H^1(G_k, M)$ is *unramified at v* if its restriction to $H^1(I_v, M)$ is trivial. (Note that I_v is defined only up to conjugation, but

the triviality or nontriviality of the restriction of ϕ is independent of the choice of I_v.)

Notation. Let M be a G_k-module, and let S be a finite set of places of k. We denote by $H^1_S(G_k, M)$ the subgroup of $H^1(G_k, M)$ consisting of cohomology classes that are unramified at all places not in S.

Proposition C.4.2. (a) *Let M be a finite G_k-module, and let S be a finite set of places of k. Then the group $H^1_S(G_k, M)$ of cohomology classes unramified outside S is finite.*
(b) *Let $\alpha : A \to B$ be an isogeny of abelian varieties over k, let S be a finite set of places of k containing:*

(i) *all archimedean places of k;*
(ii) *all places of bad reduction of A and B (in fact, one can show that A and B have bad reduction at exactly the same places);*
(iii) *all places dividing* $\deg(\alpha)$.

Then the Selmer group $\mathrm{Se}\ell^{(\alpha)}(A/k)$ is contained in $H^1_S(G_k, \ker(\alpha))$. In particular, the Selmer group if finite.

PROOF. (a) The fact that G_k acts continuously on the finite set M means that it contains an open subgroup that acts trivially. Hence there is a finite Galois extension K/k such that G_K acts trivially on M. Now the inflation-restriction sequence

$$0 \longrightarrow H^1(G_{K/k}, M^{G_{K/k}}) \xrightarrow{\text{inf}} H^1(G_k, M) \xrightarrow{\text{res}} H^1(G_K, M)$$

shows that it suffices to prove the result for K. Thus replacing k by K, we are reduced to the case that G_k acts trivially on M.

Let m be an exponent for the group M, that is, every element of M is killed by m. Then elements of

$$H^1(G_k, M) = \mathrm{Hom}(G_k, M)$$

correspond to finite abelian extensions of k whose Galois group has exponent m; and elements of $H^1_S(G_k, M)$ correspond to finite abelian extensions of k of exponent m and unramified outside of S. Hence the finiteness of $H^1_S(G_k, M)$ follows from the fact, proven above as Corollary C.1.8, that the maximal abelian extension of exponent m and unramified outside S is finite.
(b) Let $\phi \in \mathrm{Se}\ell^{(\alpha)}(A/k)$ and let v be a finite place not in S. Choose a point $y \in A(\bar{k}_v)$ such that $\phi(\sigma) = \sigma(y) - y$ for $\sigma \in G_v$. Now for σ in the inertia group I_v we compute the reduction modulo v:

$$\widetilde{\sigma(y) - y} = \widetilde{\sigma(y)} - \tilde{y} = \tilde{0}.$$

But $\sigma(y) - y$ is a torsion point (it is in $\ker(\alpha)$), and by the conditions on S, the place v is a finite place of good reduction not dividing $\deg(\alpha)$,

so Theorem C.1.4, proven in Section C.2, tells us that $\sigma(y) - y = 0$. This proves that $\phi(\sigma) = 0$ for all $\sigma \in I_v$, so ϕ is unramified at v. □

Notice that the bound for the rank of $A(k)$ obtained in Theorem C.1.9 is actually a bound for the m-rank of $H^1_S(G_k, A_m)$. Thus the present descent is a refinement of our earlier result. For further refinements of the descent, see for example Mazur [1]. The proof of the weak Mordell–Weil theorem shows that for any integer $n \geq 1$, the n-torsion subgroup $Ш(A/k)[n]$ is finite. Conjecturally, a far stronger result holds.

Conjecture C.4.3. *Let A/k be an abelian variety defined over a number field. Then $Ш(A/k)$ is finite.*

This conjecture is known to be true only for certain special elliptic curves and abelian varieties related to modular curves. Indeed, until the mid-1980s, it was not known to be true for even a single example! The first cases proven were for certain complex multiplication elliptic curves by Rubin [1] and certain modular elliptic curves by Kolyvagin [1].

C.5. Appendix: Galois Cohomology and Homogeneous Spaces

In order to motivate the technical definition of the cohomology group H^1, we discuss first, informally, the following problem:

> **Classification of Twists.** Let X_0 be an an algebro-geometric structure (an algebra, a variety, a quadratic form, etc.) defined over a field k. Classify the set of k-isomorphism classes of objects X defined over k with the property that X is isomorphic to X_0 over $\bar{k}$. The objects X are called *twists of* X_0 *over* k.

For example, if X_0/k is a curve, then the twists of X_0 are curves defined over k that are isomorphic to X_0 over $\bar{k}$. A specific example is provided by the plane cubic curves

$$X_{a,b,c} : ax^3 + by^3 + cz^3 = 0$$

for $a, b, c \in k^*$. These curves are all twists of one another, since the change of variables $[x, y, z] \mapsto [x/a^{1/3}, y/b^{1/3}, z/c^{1/3}]$ shows that they are all isomorphic (over $\bar{k}$) to $X_{1,1,1}$. However, it is a much more difficult question to determine which of the $X_{a,b,c}$'s are isomorphic over k, even for $k = \mathbb{Q}$.

Returning now to the general case, let X be a twist of X_0, and fix a $\bar{k}$-isomorphism $f : X \to X_0$. Then for each $\sigma \in G = \mathrm{Gal}(\bar{k}/k)$ we obtain an automorphism of X_0 via the composition

$$X_0 \xrightarrow{f^{-1}} X \xrightarrow{f^{\sigma}} X_0.$$

Thus we obtain a map

$$\phi : G \longrightarrow \operatorname{Aut}(X_0), \qquad \phi(\sigma) = f^\sigma \circ f^{-1}.$$

The map ϕ is not a homomorphism. A simple calculation shows that it satisfies the relation

$$\phi(\sigma\tau) = f^{\sigma\tau}f^{-1} = f^{\sigma\tau}(f^\tau)^{-1}f^\tau f^{-1} = (f^\sigma f^{-1})^\tau f^\tau f^{-1} = \phi(\sigma)^\tau \phi(\tau).$$

The map is thus a sort of "twisted homomorphism," that is, a homomorphism twisted by the action of G. Such a map is also called a *1-cocycle from G to* $\operatorname{Aut}(X_0)$.

In most situations, the 1-cocycle ϕ will have one other very important property, namely there will exist a finite Galois extension K/k such that ϕ is trivial on the subgroup $\operatorname{Gal}(\bar{k}/K)$. This is true because typical algebro-geometric objects and maps (such as varieties and rational maps between them) are determined by a finite number of polynomials that have only a finite number of coefficients. A 1-cocycle with this property is said to be *continuous*, because it is continuous with respect to the Krull topology on G and the discrete topology on $\operatorname{Aut}(X_0)$.

We also want to know to what extent the map ϕ depends on the choice of $\bar{k}$-isomorphism f. Suppose that $f_1 : X \to X_0$ is another $\bar{k}$-isomorphism, and define similarly $\phi_1(\sigma) := f_1^\sigma f_1^{-1}$. Then

$$\begin{aligned}(f_1 f^{-1})^\sigma \phi(\sigma) &= (f_1 f^{-1})^\sigma (f^\sigma f^{-1}) \\ &= f_1^\sigma f^{-1} = (f_1^\sigma f_1^{-1})(f_1 f^{-1}) = \phi_1(\sigma)(f_1 f^{-1}).\end{aligned}$$

So if we let $a := f_1 f^{-1} \in \operatorname{Aut}(X_0)$, then ϕ and ϕ_1 are related by the formula

$$a^\sigma \phi(\sigma) = \phi_1(\sigma) a \qquad \text{for all } \sigma \in G.$$

Thus a twist of X_0 leads to a 1-cocycle ϕ in the set

$$\{\phi : G \to \operatorname{Aut}(X_0) \mid \phi(\sigma\tau) = \phi(\sigma)^\tau \phi(\tau)\},$$

and the 1-cocycle ϕ is well-defined up to the equivalence relation $\phi \sim \phi_1$ if there exists an $a \in \operatorname{Aut}(X_0)$ such that $a^\sigma \phi(\sigma) = \phi_1(\sigma) a$.

Let H denote the set of equivalence classes of (continuous) 1-cocycles ϕ as above. We have just seen that each twist of X_0 gives rise to a well-defined element of H. Suppose now that X and X' are twists of X_0 that give the same element of H. This means that if we fix $\bar{k}$-isomorphisms $f : X \to X_0$ and $f' : X' \to X_0$, then there is an element $a \in \operatorname{Aut}(X_0)$ satisfying $a^\sigma \phi(\sigma) = \phi'(\sigma) a$, where $\phi(\sigma) = f^\sigma f^{-1}$ and $\phi'(\sigma) = f'^\sigma f'^{-1}$. Substituting and rearranging yields

$$(f'^{-1} a f)^\sigma = f'^{-1} a f \qquad \text{for all } \sigma \in G.$$

In other words, the isomorphism $f'^{-1}af : X \to X'$ is defined over k, so X and X' are actually the same twist of X_0. This completely formal argument shows that the map from the set of twists of X_0 to the set H is injective.

In order to show that this map is bijective, we must show that every element of H corresponds to an actual twist of X_0. This cannot be proven formally, and indeed it requires using additional structural properties of X_0. The basic idea is to take a ϕ representing an element of H, use it to construct an object X_ϕ, and show that X_ϕ is defined over k and is $\bar{k}$-isomorphic to X_0.

We illustrate this idea for a quasi-projective variety X_0/k. Let $\phi : G \to \operatorname{Aut}(X_0)$ be a continuous 1-cocycle representing an element of H. Choose a finite Galois extension K/k such that ϕ is trivial on G_K. This means that ϕ factors through $G_{K/k} := \operatorname{Gal}(K/k)$. We let $G_{K/k}$ act on $X_0 \times_k K$ by twisting the natural action via ϕ. Then the quotient $(X_0 \times_k K)/G_{K/k}$ is the desired object X_ϕ, assuming of course that this quotient object exists in the desired category. We will explain how to construct this quotient in the case that X_0 is affine; the general case then follows by a gluing argument.

So suppose that X_0/k is an affine variety, say $X_0 = \operatorname{Spec} A_0$. Thus A_0 is a k-algebra, and we let $A = A_0 \otimes_k K$ be the K-algebra obtained by extending scalars from k to K. Note that automorphisms of X_0 defined over K are in one-to-one correspondence with K-algebra automorphisms of A. In particular, each automorphism $\phi(\sigma) \in \operatorname{Aut}(X_0)$ corresponds to a K-algebra automorphism $\phi(\sigma)^* : A \to A$.

We define the ring A_ϕ to be equal to A as an abstract ring, but we twist the action of $G_{K/k}$ on A_ϕ according to the following formula:

$$\sigma \in G_{K/k} \quad \text{acts on} \quad \alpha \in A_\phi \quad \text{via} \quad \sigma \cdot \alpha := \phi(\sigma)^*(\alpha^\sigma).$$

In this definition, $\sigma \cdot \alpha$ denotes the action of σ on α considered as an element of A_ϕ, while α^σ denotes the action of σ on α considered as an element of A.

Now let $B := A_\phi^{G_{K/k}}$ be the subalgebra of A_ϕ fixed, via this action, by every element of $G_{K/k}$. One checks that B is an integral k-algebra, and Hilbert's theorem (cf. Proposition 3.1, Section A.8) tells us that B is finitely generated. Hence $X_\phi = \operatorname{Spec}(B)$ is an affine k-variety. Since $B = A \cap K(X_0)^{G_{K/k}}$ and $K \otimes k(X_0) = K(X_0)$, we see that $B \otimes_k K \cong A$; hence X_ϕ is indeed isomorphic to X_0 over K. Finally, one checks from the definition that the K-isomorphism $X \cong X_0$ is associated to the 1-cocycle ϕ.

We hope that the previous discussion has sufficiently motivated the following definitions.

Definition. Let G be a (finite or topological) group acting on another (not necessarily abelian) group A. Denote the action of G on A by $(\sigma, a) \mapsto a^\sigma$.

The 0th *cohomology group of* G *acting on* A is the group

$$H^0(G, A) = A^G = \{\alpha \in A \mid \alpha^\sigma = \alpha \text{ for all } \sigma \in G\}$$

of elements of A that are fixed by every element of G.

A map $\phi : G \to A$ is called a *1-cocycle from G to A* if it satisfies

$$\phi(\sigma\tau) = \phi(\sigma)^\tau \phi(\tau) \qquad \text{for all } \sigma, \tau \in G.$$

Two 1-cocycles ϕ, ϕ' are said to be *cohomologous* if there exists an $a \in A$ such that

$$a^\sigma \phi(\sigma) = \phi_1(\sigma) a \qquad \text{for all } \sigma \in G.$$

This is an equivalence relation, and the set of cohomology classes of 1-cocycles is denoted by $H^1(G, A)$ and is called the 1st *cohomology set of G acting on A*. If G is a topological group, for example the Galois group of an infinite extension, we will add the requirement that the cocycles should be continuous when A is given the discrete topology. For example, if A is finite, this amounts to requiring that each 1-cocycle factor through a finite quotient group of G. (Note that different cocycles may factor through different finite quotients.)

Example C.5.1. We can rephrase what we proved earlier in this section as follows: Let X_0 be a quasi-projective variety defined over k. Then there is a natural bijection between the k-twists of X_0 and the cohomology set $H^1(\operatorname{Gal}(\bar{k}/k), \operatorname{Aut}(X_0))$. (Recall that a k-twist of X_0 is a k-isomorphism class of varieties X/k such that X is $\bar{k}$-isomorphic to X_0.)

The cohomology set $H^1(G, A)$ is an example of a *pointed set* because it has a distinguished element 0, i.e., the trivial cocycle. If the group A is abelian, then it turns out that the group law on A induces a well-defined group law on $H^1(G, A)$, so in this situation $H^1(G, A)$ becomes a cohomology group, rather than merely a pointed set.

More precisely, if A is abelian, and if $\phi, \phi' : G \to A$ are 1-cocycles, we define their sum by

$$(\phi + \phi')(\sigma) = \phi(\sigma) + \phi'(\sigma).$$

It is clear from the commutativity of A that $\phi + \phi'$ is again a cocycle, so the set of cocycles forms a group, often denoted by $Z^1(G, A)$. We next define the *group of coboundaries*, denoted by $B^1(G, A)$, to be the set of maps of the form

$$\delta : G \longrightarrow A, \qquad \delta(\sigma) = a^\sigma - a,$$

where a is any element of A. One easily checks that the sum of two coboundaries is a coboundary and that every coboundary is a cocycle, so $B^1(G, A)$ is a subgroup of $Z^1(G, A)$. The cohomology group $H^1(G, A)$ is then the quotient group

$$H^1(G, A) = Z^1(G, A)/B^1(G, A).$$

The next proposition summarizes some of the basic properties of group cohomology, at least for H^0 and H^1.

Proposition C.5.2. *Let G be a group, and let A and A' be groups on which G acts.*
(1) (Functoriality) *Let $f : A \to A'$ be a G-homomorphism, that is, a homomorphism that commutes with the action of G. Then f induces a natural map*

$$H^1(G,A) \longrightarrow H^1(G,A'), \qquad [\phi] \longmapsto [f \circ \phi].$$

If A and A' are abelian, this map is a homomorphism.

Next let $F : G' \to G$ be a homomorphism. Then G' acts on A via F, and this induces a natural map

$$H^1(G,A) \longrightarrow H^1(G',A), \qquad [\phi] \longmapsto [F \circ \phi].$$

If A is abelian, this map is a homomorphism.
(2) (Inflation-restriction sequence) *Let H be a subgroup of G. Then the map $H^1(G,A) \to H^1(H,A)$ from (1) is called the* restriction map. *If further H is a normal subgroup of G, then G/H acts on A^H. In this case, the projection map $\pi : G \to G/H$ and the inclusion $A^H \hookrightarrow A$ induce the* inflation map *defined by the formula*

$$H^1(G/H,A^H) \longrightarrow H^1(G,A), \qquad [\phi] \longmapsto [\phi \circ \pi].$$

The following sequence, called the inflation-restriction sequence, *is exact:*

$$0 \longrightarrow H^1(G/H,A^H) \xrightarrow{\text{inf}} H^1(G,A) \xrightarrow{\text{res}} H^1(H,A).$$

(3) (Long exact sequence) *Let*

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

be a short exact sequence, where f and g are G-homomorphisms. Then there is a canonical long exact sequence

$$\begin{aligned} 0 \longrightarrow H^0(G,A) &\xrightarrow{f} H^0(G,B) \xrightarrow{g} H^0(G,C) \\ \xrightarrow{\delta} H^1(G,A) &\xrightarrow{f} H^1(G,B) \xrightarrow{g} H^1(G,C), \end{aligned}$$

where recall that $H^0(G,A) = A^G$, and similarly for B and C. The con- necting homomorphism *δ is defined as follows: Let $c \in H^0(G,C)$. Choose some $b \in B$ such that $g(b) = c$. Then for any $\sigma \in G$, we have*

$$g(b^\sigma - b) = g(b)^\sigma - g(b) = c^\sigma - c = 0 \quad \text{since } c \in C^G.$$

Thus $b^\sigma - b$ is in $\ker(g) = \text{Image}(f)$, and the injectivity of f means that we obtain a well-defined element $f^{-1}(b^\sigma - b) \in A$. The map

$$G \longrightarrow A, \qquad \sigma \longmapsto f^{-1}(b^\sigma - b),$$

is a cocycle representing the cohomology class of $\delta(c)$*.*

If A*,* B*, and* C *are abelian, the long exact sequence gives rise to the following useful short exact sequence:*

$$0 \longrightarrow C^G/g(B^G) \xrightarrow{\delta} H^1(G,A) \xrightarrow{f} H^1(G,B)[g] \longrightarrow 0.$$

We will not prove these basic facts about group cohomology. See, for example, Atiyah-Wall [1], Serre [4, Chapitre VII, VIII], or Silverman [1, Appendix B].

Example C.5.2.1. (Kummer sequence) Let $G_k = \mathrm{Gal}(\bar{k}/k)$ and let $\mu_m \subset \bar{k}^*$ be the group of mth-roots of unity. The short exact sequence of G_k-modules

$$0 \longrightarrow \mu_m \longrightarrow \bar{k}^* \xrightarrow{m} \bar{k}^* \longrightarrow 0$$

induces the long exact sequence

$$0 \longrightarrow \mu_m(k) \longrightarrow k^* \xrightarrow{m} k^* \xrightarrow{\delta} H^1(G_k,\mu_m) \longrightarrow H^1(G_k,\bar{k}^*).$$

Hilbert's theorem 90 says that the last group $H^1(G_k,\bar{k}^*)$ is trivial. (See Section A.2, Exercise 6, or Serre [1, Chapter 10, Proposition 2] for a proof of Hilbert's theorem 90.) This means that δ is an isomorphism, the *Kummer isomorphism*

$$k^*/k^{*m} \overset{\delta}{\cong} H^1(G,\mu_m).$$

In particular, if $\mu_m \subset k^*$, then G acts trivially on μ_m, so $k^*/k^{*m} \cong \mathrm{Hom}(G,\mu_m)$. Notice that the group $\mathrm{Hom}(G,\mu_m)$ classifies abelian extensions of k of exponent m via the association $f : G \to \mu_m$ goes to $\bar{k}^{\ker(f)}$. Further, the isomorphism $\delta : k^*/k^{*m} \to \mathrm{Hom}(G,\mu_m)$ from above is given by

$$\delta(a)(\sigma) = \alpha^\sigma/\alpha, \qquad \text{where } \alpha \text{ satisfies } \alpha^m = a.$$

Hence if $\mu_m \subset k^*$ and if K/k is an abelian extension of k of exponent m, then there always exists an element $a \in k$ such that $K = k(\sqrt[m]{a})$.

We now restrict our attention to the case of an abelian variety A defined over k. The group of automorphisms $\mathrm{Aut}(A)$ of A, considered as an abstract group with no further structure, is nonabelian. Within $\mathrm{Aut}(A)$ we can identify two important subgroups:

(i) The subgroup of translations, which we identify with $A(\bar{k})$.
(ii) The subgroup of automorphisms that fix the identity element 0, which we denote by $\mathrm{Aut}(A,0)$.

It is not hard to show that $\mathrm{Aut}(A)$ is the semidirect product $A(\bar{k}) \rtimes \mathrm{Aut}(A,0)$ (cf. Exercise A.4.15). We remark that the group $\mathrm{Aut}(A,0)$ always contains at least the two elements $[\pm 1]$. If A is an elliptic curve,

then $\operatorname{Aut}(A, 0)$ is finite, but it may be infinite for higher-dimensional abelian varieties.

The group $H^1(G_k, A(\bar{k}))$ appears naturally in the cohomological proof of the Mordell–Weil theorem, which suggests that we should look at the twists of A corresponding to the elements in this subset of $H^1(G_k, \operatorname{Aut}(A))$. It turns out these twists have a simple geometric description as homogeneous spaces.

Definition. Let A be an abelian variety defined over k. A *principal homogeneous space* of A/k is a variety X/k together with a simple transitive action of A on X also defined over k.

In other words, there is a k-morphism

$$\mu : X \times A \to X$$

with the following properties:

(i) $\mu(x, 0) = x$ for all $x \in X$.

(ii) $\mu(x, a + b) = \mu\bigl(\mu(x, a), b\bigr)$ for all $x \in X$ and all $a, b \in A$.

(iii) For all $x \in A$, the map $a \mapsto \mu(x, a)$ is an isomorphism $A \to X$ (defined over $k(x)$).

In particular, (iii) tells us that X is a twist of A, since it is isomorphic to A over $\bar{k}$. Further, it is clear that X is isomorphic to A over k if and only if $X(k) \neq \emptyset$. Thus if $x \in X(k)$, then the map in (iii) gives a k-isomorphism. Conversely, if X is k-isomorphic to A, then this isomorphism maps the point $0 \in A(k)$ to a k-rational point of X.

Two homogeneous spaces (X, μ) and (X', μ') for A/k are isomorphic (as homogeneous spaces) if there is a k-isomorphism $i : X \to X'$ such that the following diagram commutes:

$$\begin{array}{ccc} X \times A & \xrightarrow{\mu} & X \\ \big\downarrow{\scriptstyle i\times 1} & & \big\downarrow{\scriptstyle i} \\ X' \times A & \xrightarrow{\mu'} & X' \end{array}$$

In other words, the isomorphism $i : X \to X'$ is required to commute with the action of A on X and X'. Note that it is possible for a twist X/k of A/k to have several nonisomorphic structures as a homogeneous space of A. (See, e.g., Silverman [1, Exercise 10.4].)

The fact that the action μ is principal allows us to define a "subtraction map" $\nu : X \times X \to A$ as follows: $\nu(x, y)$ is the unique point of A/k satisfying

$$\mu\bigl(y, \nu(x, y)\bigr) = x.$$

The existence and uniqueness of the point $\nu(x, y)$ follows from property (iii). An alternative definition of ν is to fix any point $x_0 \in X$, let $\theta : A \to X$ be the isomorphism $\mu(x_0, \cdot\,)$ given by (iii), and then define $\nu(x, y) = \theta^{-1}(x) - \theta^{-1}(y)$. We leave it to the reader to verify that ν is a well-defined k-morphism.

Proposition C.5.3. *There is a natural bijection between (k-isomorphism classes of) principal homogeneous spaces of A/k and $H^1(G_k, A(\bar{k}))$. This association is given as follows: Let X be a principal homogeneous space of A/k, choose any point $x \in X$, and let $\nu : X \times X \to A$ be the subtraction map defined above. Then*

$$[X] \longmapsto [\sigma \mapsto \nu(\sigma(x), x)].$$

PROOF. See Silverman [1, Theorem 3.6], where the proof given for elliptic curves applies without change to abelian varieties. We also note that choosing a different point in X merely changes the cocycle by a coboundary. Thus if $x' \in X$, then the difference of the cocycles corresponding to x and x' is the coboundary

$$\nu(\sigma(x'), x') - \nu(\sigma(x), x) = \sigma(\nu(x', x)) - \nu(x', x).$$

□

Proposition C.5.3 tells us the somewhat surprising fact that the set of principal homogeneous spaces of an abelian variety A has a natural group structure. This group is called the *Weil–Châtelet group of A/k* and is denoted by $\mathrm{WC}(A/k)$. The group law on $\mathrm{WC}(A/k)$ can be (and historically was first) defined geometrically. This geometric construction is described in Exercise C.17.

EXERCISES

C.1. Let G be an abelian group, let $m \geq 2$ an integer such that the quotient G/mG is finite, and let $x_1, \ldots, x_s \in G$ be a complete set of coset representatives for G/mG. Suppose that there are constants $A, B, C, D \geq 0$ with $A > B$ (depending on G, m, and $x_1, \ldots, x_s$) and a function $h : G \to \mathbb{R}$ with the property that

$$h(mx) \geq A\big(h(x) - C\big) \quad \text{and} \quad h(x + x_i) \leq Bh(x) + D$$

for all $x \in G$ and $1 \leq i \leq s$. Prove that the set

$$\left\{ x \in G \;\middle|\; h(x) \leq \frac{C + D}{A - B} \right\}$$

generates the group G.

C.2. Let A and B be abelian varieties defined over a number field k, let v be a place of k at which A and B both have good reduction, and let $\tilde{A}_v$ and $\tilde{B}_v$ denote the reductions. Show that

$$\mathrm{Hom}(A, B) \longrightarrow \mathrm{Hom}(\tilde{A}_v, \tilde{B}_v)$$

is injective. (*Hint.* Use Theorem C.1.4 to show that if $\Phi \neq 0$, then $\tilde{\Phi}$ cannot vanish on all torsion points.)

C.3. Give a bound, or even better compute exactly, the quantity $\operatorname{card}(A_{\text{tor}}(\mathbb{Q}))$ for the following elliptic curves $A/\mathbb{Q}$:
(a) $y^2 = x^3 - 1$.
(b) $y^2 = x^3 - 4x$.
(c) $y^2 = x^3 + 4x$.
(d) $y^2 + 17xy - 120y = x^3 - 60x^2$.

C.4. (Weil-style method of computing the cardinality of $\operatorname{Jac}(C)(\mathbb{F}_p)$) Let C be a curve of genus g defined over $\mathbb{F}_p$, and let $J = \operatorname{Jac}(C)$ be its Jacobian variety. For each integer $m \geq 1$, let

$$N_m(C) = \operatorname{card} C(\mathbb{F}_{p^m}) \qquad \text{and} \qquad N_m(J) = \operatorname{card} J(\mathbb{F}_{p^m}).$$

We know from Exercise A.8.11 that there exist algebraic integers α_i such that

$$N_m(C) = p^m + 1 - (\alpha_1^m + \cdots + \alpha_{2g}^m) \quad \text{for all } m \geq 1.$$

Furthermore, the polynomial $P(T) := \prod_{i=1}^{2g}(1 - \alpha_i T)$ has integer coefficients and leading coefficient p^g, and it satisfies $P(T) = p^g T^{2g} P(1/pT)$. Then

$$N_1(J) = \operatorname{card} J(\mathbb{F}_p) = P(1) = \prod_{i=1}^{2g}(1 - \alpha_i).$$

Prove that the first g cardinalities $N_1(C), N_2(C), \ldots, N_g(C)$ for C determine the cardinality $N_1(J)$. In particular, prove that when $g = 2$,

$$N_1(J) = \frac{1}{2}(N_1(C)^2 + N_2(C)) - p.$$

Find a similar formula for $g = 3$. (*Hint.* Use Newton's formulas relating elementary symmetric polynomials to sums of powers.)

C.5. Let A be the Jacobian of the curve $y^2 = x^5 - x$. Compute the torsion subgroup $A_{\text{tors}}(\mathbb{Q})$. (*Hint.* Use Exercise A.8.1 to determine the rational 2-torsion points in $A(\mathbb{Q})$. Then use the previous exercise and reduce modulo 3 and modulo 5 to prove that $A_{\text{tors}}(\mathbb{Q})$ is generated by its 2-torsion and possibly a single rational 3-torsion point. Finally, determine whether or not there is such a 3-torsion point.)

C.6. Let p be an odd prime, let r and s be integers satisfying $0 < r, s, r + s < p$, and let C be the smooth projective curve birational to $y^p = x^r(x-1)^s$. In this exercise you will prove that $\operatorname{Jac}(C)(\mathbb{Q})_{\text{tor}}$ is isomorphic to either $\mathbb{Z}/2p\mathbb{Z}$ or $\mathbb{Z}/p\mathbb{Z}$.
(a) Show that the quasi-affine curve defined by

$$U = \{(x, y) \in \mathbb{A}^2 \mid y^p = x^r(x-1)^s \text{ and } x(x-1) \neq 0\}$$

is smooth, and hence that U is an open subset of C. Prove that the complement $C \smallsetminus U = \{P_0, P_1, P_\infty\}$ consists of exactly three points, where P_0, P_1, P_∞ are the points above $(0,0)$, $(1,0)$, and ∞, respectively. Prove that the genus g of C is equal to $(p-1)/2$.

(b) Show that

$$\begin{aligned}\operatorname{div}(x) &= p(P_0) - p(P_\infty),\\ \operatorname{div}(x-1) &= p(P_1) - p(P_\infty),\\ \operatorname{div}(y) &= r(P_0) + s(P_1) - (r+s)(P_\infty).\end{aligned}$$

(c) Suppose that p does not divide $q-1$. Prove that $\operatorname{card}(C(\mathbb{F}_q)) = q+1$. Use Exercise C.4 to deduce that if ℓ is a primitive root modulo p, then $\operatorname{card}(\operatorname{Jac}(C)(\mathbb{F}_\ell)) = \ell^g + 1$.
(d) Prove that $\operatorname{Jac}(C)(\mathbb{Q})_{\mathrm{tor}}$ is isomorphic to either $\mathbb{Z}/2p\mathbb{Z}$ or $\mathbb{Z}/p\mathbb{Z}$. (*Hint.* Use Dirichlet's theorem on primes in arithmetic progressions, which says that if $\gcd(a,b) = 1$, then there are infinitely many primes of the form $an+b$.)
(e) Prove that in fact $\operatorname{Jac}(C)(\mathbb{Q})_{\mathrm{tor}} \cong \mathbb{Z}/p\mathbb{Z}$ except in the one case that $p = 7$ and $r^3 \equiv s^3 \equiv -(r+s)^3 \pmod 7$. (This is more difficult. See Gross–Rohrlich [1] for details).

C.7. In this exercise you will prove the Chevalley–Weil theorem.

Chevalley–Weil Theorem. *Let $\phi : X \to Y$ be an unramified covering of normal projective varieties defined over a number field k. Then there exists a finite extension K/k such that $\phi^{-1}(Y(k)) \subset X(K)$.*

Before beginning, we make a definition. An *M_k-constant* is a map

$$\gamma : M_k \longrightarrow \mathbb{R}$$

with the properties (i) $\gamma(v) > 0$ for all $v \in M_k$ and (ii) $\gamma(v) = 1$ for all but finitely many $v \in M_k$. (Note that the M_k-constants defined in Section B.8 are the logarithms of these M_k-constants.)
(a) Let U be an affine variety with coordinate ring $k[U] = k[f_1,\dots,f_m]$. For each place v of k, let

$$U_v = \{x \in U(k) \mid v(f_i(x)) \geq 0 \text{ for all } 1 \leq i \leq m\}.$$

Informally, we say that U_v is the set of v-integral points of U. (Of course, U_v depends on the choice of the f_i's.) Let $g \in k[U]^*$. Prove that there are M_k-constants γ_1, γ_2, depending on U, g, and $f_1,\dots,f_m$, such that

$$\gamma_1(v) \leq |g(x)|_v \leq \gamma_2(v) \quad \text{for all } v \in M_k \text{ and all } x \in U_v.$$

(b) Let U/k and V/k be affine varieties, and let $\phi : V \to U$ be a morphism with the property that $k[V]$ is a free $k[U]$ module of rank n, and further suppose that there exists a basis $g_1,\dots,g_n$ for $k[V]/k[U]$ whose discriminant is a unit in $k[U]$, i.e.,

$$\operatorname{Disc}_{k[U]}^{k[V]}(g_1,\dots,g_n) \in k[U]^*.$$

Prove that there exist M_k-constants γ_3, γ_4, depending on all of the above data, such that

$$\gamma_3(v) \leq \left|\operatorname{Disc}(k(\phi^{-1}(x))/k)\right|_v \leq \gamma_4(v) \quad \text{for all } v \in M_k \text{ and all } x \in U_v.$$

(c) Now assume that U/k and V/k are normal affine varieties, and let $\phi : V \to U$ be an unramified covering. Let

$$U(R_k) := \{x \in U(k) \mid f_i(x) \in R_k \text{ for all } 1 \le i \le m\}$$

be the set of integral points of U. (Again, this set clearly depends on the choice of coordinate functions $f_1, \dots, f_m$ on U.) Prove that there is a constant c_1, depending on these data, such that

$$|\mathrm{Disc}(k(\phi^{-1}(x))/k)| \le c_1 \quad \text{for all } x \in U(R_k).$$

(*Hint.* Cover U with affine open subsets to which you can apply part (b).)
(d) Finally, suppose that U/k and V/k are normal projective varieties, and again let $\phi : V \to U$ be an unramified covering. Prove that there is a constant c_2, depending on k, U, V, and ϕ, such that

$$|\mathrm{Disc}(k(\phi^{-1}(x))/k)| \le c_2 \quad \text{for all } x \in U(k).$$

(*Hint.* Cover U with affine open subsets to which you can apply part (c).)
(e) Under the hypotheses of (c), prove that there is a finite extension K/k such that $\phi^{-1}(U(R_k)) \subset V(K)$. Similarly, under the hypotheses of (d), prove that there is a finite extension K/k such that $\phi^{-1}(U(k)) \subset V(K)$. (*Hint.* Use Hermite's theorem (C.3.2).)

C.8. Let k be a number field, and let $K = k[\beta]$ be a finite extension of k generated by the algebraic integer β. Let $F(X) \in R_k[X]$ be the minimal polynomial of β.
(a) Prove that the discriminant of the order $R_k[\beta]$ over R_k is equal to $\pm N_k^K(F'(\beta))$.
(b) Prove that the only primes that can ramify in the extension K/k are the primes dividing $N_k^K(F'(\beta))$.
(c) Let $\alpha \in k$ be an algebraic integer. Prove that the extension $k(\sqrt[m]{\alpha})/k$ is unramified except possibly at primes dividing $m\alpha$.

C.9. *Hensel's lemma and an application to torsion points.*
Let k be a p-adic field, i.e., the completion of a number field with respect to a nonarchimedean place, let R be the ring of integers of k, and let π be a uniformizer (a generator of the maximal ideal).
(a) Let $P \in R[X]$, and let $x_0 \in R$ be an element satisfying

$$P(x_0) \equiv 0 \pmod{\pi} \qquad \text{and} \qquad P'(x_0) \not\equiv 0 \pmod{\pi}.$$

Prove that there exists a unique $x \in R$ satisfying

$$P(x) = 0 \qquad \text{and} \qquad x \equiv x_0 \pmod{\pi}.$$

This result is the classical version of *Hensel's lemma.* (*Hint.* Construct x as the limit of a sequence $x_0, x_1, x_2, \dots$ satisfying

$$P(x_m) \equiv 0 \pmod{\pi^{m+1}} \qquad \text{and} \qquad x_m \equiv x_{m-1} \pmod{\pi^m}.)$$

(b) Generalize part (a) as follows. Let $P_1, \dots, P_r \in R[X_1, \dots, X_s]$ be a collection of polynomials, and let $x_0 \in R^s$ be a point such that

$$P_1(x_0) \equiv \cdots \equiv P_r(x_0) \equiv 0 \,(\mathrm{mod}\ \pi),$$

and such that the matrix

$$\left(\frac{\partial P_i}{\partial X_j}(x_0) \,(\mathrm{mod}\ \pi)\right)_{\substack{1 \le i \le r \\ 1 \le j \le s}}$$

has rank r. Prove that there exists a point $x \in R^s$ satisfying

$$P_1(x) = \cdots = P_r(x) = 0 \qquad \text{and} \qquad x \equiv x_0 \pmod{\pi}.$$

(c) Let Y be a variety defined over k, and let $\tilde{Y}$ be the reduction $(\mathrm{mod}\ \pi)$. Let $\tilde{P} \in \tilde{Y}(R/\pi)$ be a nonsingular point of $\tilde{Y}$. Prove that there is a point $P \in Y(k)$ whose reduction modulo π is equal to $\tilde{P}$. In particular, if $\tilde{Y}(R/\pi)$ contains a nonsingular point, then $Y(k)$ is nonempty.
(d) Let A be an abelian variety defined over k with good reduction at π. Let m be an integer not divisible by π, and assume that all of the m-torsion of A is k-rational (i.e., $A_m \subset A(k)$). Prove that the reduction map $A_m \to \tilde{A}_m$ is onto and, using Theorem A.7.2.7, conclude that $A_m \cong \tilde{A}_m$. (Note that $\tilde{A}_m$ denotes the m-torsion on $\tilde{A}$, not the reduction of the m-torsion of A.)
(e) Use the results of this exercise to give another proof of Theorem C.1.4.

C.10. This exercise provides a scheme-theoretic proof of Theorem C.1.4. It requires knowledge of some nontrivial scheme theory. Let $m \ge 2$, let A be an abelian variety defined over a number field k, and let S be a finite set of places of k containing all places of bad reduction of A and all places dividing m.
(a) Show that there exists an abelian scheme $\mathcal{A} \to \mathrm{Spec}(R_S)$ with generic fiber A/k. That is, $\mathcal{A}$ is a group scheme over $\mathrm{Spec}(R_S)$ such that the fiber over every closed point of $\mathrm{Spec}(R_S)$ is an abelian variety, and such that the fiber over the generic point is A/k. Prove further that there is a $\mathrm{Spec}(R_S)$-morphism $[m] : \mathcal{A} \to \mathcal{A}$ that induces multiplication-by-m on every fiber.
(b) Let G be a subgroup of $A[m](k)$, and let v be a place of R_S. Prove that the reduction map $G \to \mathcal{A}_v(\tilde{k}_v)$ is injective. (*Hint.* Use the fact that if B is an abelian variety of dimension g, and if m is relatively prime to the characteristic of the base field, then $B[m] \cong (\mathbb{Z}/m\mathbb{Z})^{2g}$.)

C.11. *Inversion of formal power series.*
This exercise sketches a proof of Lemma C.2.2. Let R denote a commutative ring.
(a) Let $F(T) = aT + \cdots \in R[\![T]\!]$ be a formal series with $a \in R^*$. Prove then there exists a unique formal power series $G(T) \in R[\![T]\!]$ satisfying

$$G(T) = a^{-1}T + \cdots \qquad \text{and} \qquad F(G(T)) = G(F(T)) = T.$$

(*Hint.* : Construct G as the limit of a sequence of polynomials $G_0, G_1, \ldots$ satisfying $F(G_n(T)) \equiv T \pmod{T^{n+1}}$ and $G_{n+1}(T) \equiv G_n(T) \pmod{T^{n+1}}$. Notice the similarity to Hensel's lemma (Exercise C.9).)
(b) generalize (a) to the several-variables setting. Let $F = (F_1, \ldots, F_m)$ be an m-tuple of formal series in m variables, and write

$$F_i = a_{i1}T_1 + \cdots + a_{im}T_m + \cdots \in R[\![T_1, \ldots, T_m]\!].$$

Assume that the Jacobian determinant $\det(a_{ij})$ is in R^*. Prove that there exists a unique m-tuple $G = (G_1, \ldots, G_m)$ of formal series $G_i \in R[\![T_1, \ldots, T_m]\!]$ with no constant terms such that for all $1 \le i, j \le m$,

$$F_i(G_1(T), \ldots, G_m(T)) = T_i \qquad \text{and} \qquad G_j(F_1(T), \ldots, F_m(T)) = T_j.$$

C.12. For each of the following curves $C/\mathbb{Q}$, let $J = \operatorname{Jac}(C)$ and find as accurate a bound as you can for the Mordell–Weil rank of J.
(a) Let $C : y^2 = x^5 - x$. Find bounds for $\operatorname{rank} J(\mathbb{Q})$ and $\operatorname{rank} J(\mathbb{Q}(i))$. (*Hint.* Use Theorem C.1.9 and show that $\operatorname{rank} J(\mathbb{Q}(i)) = 2 \operatorname{rank} J(\mathbb{Q})$.)
(b) Let $C : y^2 = x^6 - 1$, and let $\eta = e^{2\pi i/3}$ be a primitive cube root of unity. Find bounds for $\operatorname{rank} J(\mathbb{Q})$ and $\operatorname{rank} J(\mathbb{Q}(\eta))$. (*Hint.* Use Theorem C.1.9 and show that $\operatorname{rank} J(\mathbb{Q}(\eta)) = 2 \operatorname{rank} J(\mathbb{Q})$.)
(c) Let $C : y^2 = x(x^2 - 1)(x^2 - 4)$. Find a bound for $\operatorname{rank} J(\mathbb{Q})$.

C.13. Let $P(x) \in \mathbb{Q}[x]$ be a polynomial with simple roots, let p be a prime number, and let $K = \mathbb{Q}(\exp(2\pi i/p))$. Let C be the smooth projective curve birational to the affine curve $y^p = P(x)$, and let $J = \operatorname{Jac}(C)$. Prove that $\operatorname{rank}(J(K)) = (p-1) \operatorname{rank} J(\mathbb{Q})$.

C.14. Let p be a prime number, let $n \ge 1$ be an integer, and let

$$\rho_p : \mathrm{GL}(n, \mathbb{Z}) \longrightarrow \mathrm{GL}(n, \mathbb{F}_p)$$

be the reduction modulo p map. Prove that $\ker(\rho_p)$ is trivial if $p \ge 3$, and consists of the elements of order 2 in $\mathrm{GL}(n, \mathbb{Z})$ when $p = 2$.

C.15. Let $C/\mathbb{Q}$ be the smooth projective curve birational to the affine curve $2y^2 = x^4 - 17$. This exercise sketches a proof that $C(\mathbb{Q}_v) \ne \emptyset$ for all places v of $\mathbb{Q}$, yet $C(\mathbb{Q}) = \emptyset$.
(a) Show that C has good reduction at all primes except 2 and 17, and that $\tilde{C}(\mathbb{F}_p)$ contains a nonsingular point for every prime p. Conclude that $C(\mathbb{Q}_p) \ne \emptyset$ for all primes p. (*Hint.* Use Weil's estimate (Exercise C.4) to get points modulo p, and then Hensel's lemma (Exercise C.9) to lift them to p-adic points.)
(b) Check that $C(\mathbb{R}) \ne \emptyset$.
(c) Show that the two points at infinity on C are not rational over $\mathbb{Q}$.
(d) Suppose that $C(\mathbb{Q})$ contained a point. Prove that there would then exist coprime integers a, b, c satisfying $a^4 - 17b^4 = 2c^2$.
(e) Let a, b, c be as in (d). Prove that c is a square modulo 17. (*Hint.* For odd p dividing c, use the fact that p is a square modulo 17 if and only if 17 is a square modulo p.) Conclude that 2 is a 4th power modulo 17. This contradiction implies that $C(\mathbb{Q}) = \emptyset$.

C.16. Let p be an odd prime, let r and s be integers satisfying $0 < r, s, r+s < p$, and let $C = C_{r,s}$ be the smooth projective curve birational to $v^p = u^r(1-u)^s$ that was studied in Exercise C.6. Recall that C has one point Q_0 lying over $(0,0)$ and one point Q_∞ at infinity. Also let F be the Fermat curve $F : x^p + y^p = 1$.

(a) Let ϕ be the map

$$\phi : F \longrightarrow C, \qquad (x,y) \longmapsto (x^p, x^r y^s).$$

Show that ϕ is a well-defined unramified Galois covering of degree p.

(b) Let $\zeta = \exp(2i\pi/p)$, and define maps $\alpha(x,y) = (\zeta x, y)$ and $\beta(x,y) = (x, \zeta y)$. Note that $\alpha, \beta \in \operatorname{Aut}(F)$. Let $u \geq 1$ be an integer satisfying $ru \equiv -s \pmod{p}$. Prove that $\alpha^u\beta$ generates the Galois group of the covering ϕ.

(c) Suppose that $rs' \equiv r's \pmod{p}$. Prove that $C_{r,s}$ is isomorphic to $C_{r',s'}$. (*Hint.* Write $(r', s') = k(r,s) + p(i,j)$ and consider the map $(u,v) \mapsto (u, v^k u^i (1-u)^j$.)

(d) Let D_∞ be the divisor at infinity on F (i.e., the sum of the p points on $x^p + y^p = z^p$ where $z = 0$). Prove that for any $d \geq 1$, the set of functions

$$\{x^m y^n \mid 0 \leq m, \quad 0 \leq n \leq p-1, \quad 0 \leq m+n \leq d\}$$

is a basis of $L(dD_\infty)$.

(e) Let $P_1, \ldots, P_d \in F$, and let $\gamma = \alpha^u\beta$ with $d < u < p/d$. Prove that a divisor of the form

$$D = \left(\sum_{i=1}^{d}\sum_{j=0}^{p-1}(\gamma^j P_i)\right) - dD_\infty$$

cannot be principal unless all the P_i's have one their coordinates equal to 0. (*Hint.* Suppose that $D = \operatorname{div}(f)$ with $f = \sum a_{mn}x^m y^n$. Use the invariance of D by γ to show that $f \circ \gamma = \zeta^k f$ for some k, and deduce that f is a monomial in x, y.)

(f) Let $\eta = e^{2\pi i/6}$ be a primitive 6th root of unity. Let $P = (\eta, \bar\eta)$ and $\bar P = (\bar\eta, \eta)$, and verify that $P, \bar P \in F$.

(g) Let $P, \bar P$ be as in (f), and let $Q = \phi(P)$ and $\bar Q = \phi(\bar P)$ be their images in C. Assume that either $3p/4 < s < p-4$ or $3 < s < p/4 - 1$. Prove that the divisor class of $(Q) + (\bar Q) - 2(Q_\infty)$ is a point of infinite order in $\operatorname{Jac}(C_{1,s})(\mathbb{Q})$. (*Hint.* If it were a torsion point, Exercise C.6 would say that the divisor class of $2(Q) + 2(\bar Q) - 4(Q_\infty)$ must be a multiple of the class of $(Q_0) - (Q_\infty)$. Show that this is not possible.)

(g) Assume that $2 < s < p-2$ and that $s \neq (p-1)/2$. Suppose further that $\operatorname{Jac}(C)(\mathbb{Q})_{\text{tor}} \cong \mathbb{Z}/p\mathbb{Z}$. (This is, in fact, true if $p > 7$.) Prove that the divisor class of $(Q) + (\bar Q) - 2(Q_\infty)$ is a point of infinite order in $\operatorname{Jac}(C_{1,s})(\mathbb{Q})$.

C.17. **Geometric Group Law on** WC(A/k).
Let A be an abelian variety defined over a perfect field k, and let Y_1/k and Y_2/k be two principal homogeneous spaces for A/k. Denote the action of A on Y_i by

$$Y_i \times A \longrightarrow Y_i, \qquad (y, a) \longmapsto y + a.$$

We also recall that for any $y_0 \in Y_i(\bar{k})$, the map $a \mapsto y_0 + a$ is a $\bar{k}$-isomorphism $A \xrightarrow{\sim} Y_i$.
(a) Prove that there exists a homogeneous space Y_3/k for A/k and a k-morphism $f : Y_1 \times Y_2 \to Y_3$ satisfying

$$f(y_1 + a_1, y_2 + a_2) = f(y_1, y_2) + a_1 + a_2 \qquad \text{for all } (y_1, y_2) \in Y_1 \times Y_2 \text{ and all } a_1, a_2 \in A.$$

(b) Prove that Y_3 is unique up to k-isomorphism of homogeneous spaces.
(c) Prove that Y_3 represents the sum of Y_1 and Y_2 as elements of the cohomology group $H^1(G_k, A(\bar{k}))$ (see Proposition C.5.3). In other words, the map $([Y_1], [Y_2]) \mapsto [Y_3]$ on k-isomorphism classes of homogeneous spaces coincides with the group law on $H^1(G_k, A(\bar{k}))$.

C.18. Let $f(x) = \prod_{i=1}^{2g+1}(x - \alpha_i) \in k[x]$ be a polynomial with distinct roots, and let C be the hyperelliptic curve birational to $y^2 = f(x)$. Note that C has a single point ∞ at infinity. Let $P_i = (\alpha_i, 0)$ for $1 \le i \le 2g + 1$, and let

$$\mathcal{W} = \{P_1, P_2, \ldots, P_{2g+1}, \infty\}.$$

(The set $\mathcal{W}$ is the set of Weierstrass points of C.) Let

$$\operatorname{Div}^0_{\mathcal{W}}(C) = \left\{ D = \sum n_i(Q_i) \in \operatorname{Div}(C) \;\middle|\; \deg(D) = 0 \text{ and } Q_i \notin \mathcal{W} \right\}.$$

In other words, $\operatorname{Div}^0_{\mathcal{W}}(C)$ is the set of divisors of degree 0 with support disjoint from $\mathcal{W}$. Finally, let $J = \operatorname{Jac}(C)$, let $L = k[T]/(f(T))$, and let $A = L^*/L^{*2}$.
(a) Define a map

$$\Phi : \operatorname{Div}^0_{\mathcal{W}}(C) \longrightarrow A, \qquad \sum n_i(Q_i) \longmapsto \prod (x(Q_i) - T)^{n_i}.$$

Prove that if $D, D' \in \operatorname{Div}^0_{\mathcal{W}}(C)$ are linearly equivalent, then $\Phi(D) = \Phi(D')$. (*Hint.* Use Weil's reciprocity law, Exercise A.4.16.)
(b) Show that every divisor $D \in \operatorname{Div}(C)$ of degree zero is linearly equivalent to a divisor in $\operatorname{Div}^0_{\mathcal{W}}(C)$. Use this and (a) to show that Φ induces a well-defined homomorphism $\Phi : J(k) \to A$.
(c) Prove that the kernel of $\Phi : J(k) \to A$ is equal to $2J(k)$.
(d) Prove that the image of $\Phi : J(k) \to A$ is contained in the kernel of the norm map $L^*/L^{*2} \xrightarrow{\mathrm{N}} k^*/k^{*2}$.
(e) Assume now that k is a number field, and let S be a set of places containing the places over 2 and the places of bad reduction of J. Let $A(2, S)$ denote the subgroup of elements of $A = L^*/L^{*2}$ whose square roots generate extensions of k that are unramified outside S. Prove that the image of $\Phi : J(k) \to A$ is contained in $A(2, S)$.

C.19. Let C be a smooth projective model of the curve $y^2 = x(x-2)(x-3)(x-4)(x-5)(x-7)(x-10)$. Use the previous exercise to show that the points on $\mathrm{Jac}(C)$ corresponding to the divisor classes of $(1,36)-(\infty)$ and $(6,24)-(\infty)$ are two independent points of infinite order. Conclude that $\mathrm{Jac}(C)(\mathbb{Q})$ contains a subgroup isomorphic to $(\mathbb{Z}/2\mathbb{Z})^6 \times \mathbb{Z}^2$. (In fact, one can show that $\mathrm{Jac}(C)(\mathbb{Q})$ has rank exactly 2; see Schaefer [1].)

PART D

Diophantine Approximation and Integral Points on Curves

He was a poet and hated the approximate.
R. M. Rilke, *The Journal of My Other Self*

The fundamental problem in the subject of Diophantine approximation is the question of how closely an irrational number can be approximated by a rational number. For example, if $\alpha \in \mathbb{R}$ is any given real number, we may ask how closely can one approximate α by a rational number $p/q \in \mathbb{Q}$? The obvious answer is that the difference $|(p/q) - \alpha|$ can be made as small as desired by an appropriate choice of p/q. This is nothing more than the assertion that $\mathbb{Q}$ is dense in $\mathbb{R}$. The problem is to show that if the difference is small, then p and q must be large.

More precisely, let $\alpha \in \mathbb{R}$ be a given real number, and let $e > 0$ be a given exponent. We ask whether or not the inequality

$$\left|\frac{p}{q} - \alpha\right| \leq \frac{1}{q^e}$$

can have infinitely many solutions in rational numbers $p/q \in \mathbb{Q}$. For example, a theorem of Dirichlet says that the inequality

$$\left|\frac{p}{q} - \alpha\right| \leq \frac{1}{q^2}$$

always has infinitely many solutions, while a result of Liouville says that if α is an algebraic number of degree d over $\mathbb{Q}$, and if $e > d$, then the inequality

$$\left|\frac{p}{q} - \alpha\right| \leq \frac{1}{q^e}$$

has only finitely many solutions. We will prove these elementary results of Dirichlet and Liouville in Section D.1.

In general, the *approximation exponent* of a real number $\alpha \in \mathbb{R}$ is defined to be the smallest number $\tau(\alpha)$ with the property that for any exponent $e > \tau(\alpha)$, the inequality

$$\left|\frac{p}{q} - \alpha\right| \leq \frac{1}{q^e}$$

has only finitely many solutions in rational numbers $p/q \in \mathbb{Q}$. Thus Dirichlet's theorem says that $\tau(\alpha) \geq 2$ for every real number α, and Liouville's theorem says that if α is an algebraic number of degree d, then $\tau(\alpha) \leq d$.

The exponent in Liouville's theorem for algebraic numbers of degree d has been successively improved by a number of mathematicians, as indicated in the following brief table:

Liouville	1844	$\tau(\alpha) \leq d$
Thue	1909	$\tau(\alpha) \leq \frac{1}{2}d + 1$
Siegel	1921	$\tau(\alpha) \leq 2\sqrt{d}$
Gelfand, Dyson	1947	$\tau(\alpha) \leq \sqrt{2d}$
Roth	1955	$\tau(\alpha) = 2$

Thus Roth's result can be stated in the following way:

Roth's Theorem. For every algebraic number α and every $\varepsilon > 0$, the inequality

$$\left|\frac{p}{q} - \alpha\right| \leq \frac{1}{q^{2+\varepsilon}}$$

has only finitely many rational solutions $p/q \in \mathbb{Q}$.

Equivalently, for every $\varepsilon > 0$ there exists a constant $C = C(\alpha, \varepsilon) > 0$ such that for all $p/q \in \mathbb{Q}$,

$$\left|\frac{p}{q} - \alpha\right| \geq \frac{C(\alpha, \varepsilon)}{q^{2+\varepsilon}}.$$

Roth's theorem has been extended in various ways, such as allowing several (possibly nonarchimedean) absolute values and taking approximating values from a number field K rather than from $\mathbb{Q}$. Our main goal in this chapter is to prove a general version of Roth's theorem. We will also give two important Diophantine applications. We will show that the equation $u + v = 1$ has only finitely many solutions in S-units of a number field, and we will prove Siegel's theorem, which says that an affine piece of a curve of genus at least one has only finitely many S-integer points.

D.1. Two Elementary Results on Diophantine Approximation

In this section we prove two elementary results that illustrate some of the techniques used in the study of Diophantine approximation. The first result says that we can find rational numbers that are fairly close to a given real number. It shows, in particular, that the exponent $2 + \varepsilon$ in Roth's theorem is essentially best possible.

Proposition D.1.1. (Dirichlet 1842) *Let $\alpha \in \mathbb{R}$ with $\alpha \notin \mathbb{Q}$. Then there are infinitely many rational numbers $p/q \in \mathbb{Q}$ satisfying*

$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{q^2}.$$

Remark D.1.1.1. The proof that we will give does not provide an efficient method of constructing good approximations to a given real number. The classical theory of *continued fractions* provides such a method (see Exercise D.18). A classical result of Hurwitz says that the $1/q^2$ in Dirichlet's estimate (D.1.1) may be replaced by $1/\sqrt{5}q^2$, and that this is best possible. For further information, see Exercise D.3 and the references cited there.

PROOF (of Proposition D.1.1). For any integer $Q \ge 1$, consider the set of real numbers

$$\{q\alpha - [q\alpha] \mid q = 0, 1, \dots, Q\},$$

where $[t]$ means the greatest integer in t. Since α is irrational, this set consist of $Q + 1$ distinct numbers in the interval between 0 and 1. If we divide the unit interval into Q line segments of equal length, the pigeonhole principle tells us that one of the segments must contain two of the numbers. In other words, the set contains two numbers whose distance from one another is at most $1/Q$, so we can find integers $0 \le q_1 < q_2 \le Q$ satisfying

$$\left|\left(q_1\alpha - [q_1\alpha]\right) - \left(q_2\alpha - [q_2\alpha]\right)\right| \le \frac{1}{Q}.$$

A little algebra and the estimate $1 \le q_2 - q_1 \le Q$ gives

$$\left|\frac{[q_2\alpha] - [q_1\alpha]}{q_2 - q_1} - \alpha\right| \le \frac{1}{(q_2 - q_1)Q} \le \frac{1}{(q_2 - q_1)^2}.$$

Thus for each Q we obtain a rational approximation of α with the desired property. Further, by increasing Q we can make the left-hand side as close to 0 as we wish, which means that we obtain infinitely many distinct rational approximations. □

Dirichlet's theorem (Proposition D.1.1) tells us that we can always approximate α by a rational number p/q to within $1/q^2$. The next result, due to Liouville, gives an estimate in the other direction.

Proposition D.1.2. (Liouville 1844) *Let $\alpha \in \bar{\mathbb{Q}}$ be an algebraic number of degree $d = [\mathbb{Q}(\alpha) : \mathbb{Q}] \ge 2$. Fix a constant $\varepsilon > 0$. Then there are only finitely many rational numbers $p/q \in \mathbb{Q}$ satisfying*

$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{q^{d+\varepsilon}}. \qquad (*)$$

PROOF. Although the proof of Liouville's result is very elementary, it does include many of the important features that will reappear in the proof of Roth's theorem. To emphasize the similarity, we have broken the proof into several steps. The reader might compare Steps I, II, III, and IV with the material in Sections 4, 5, 6, and 7 of this chapter.

Step I: Construction of a Polynomial

The first step is to construct a polynomial that vanishes at α. An obvious choice is to take the minimal polynomial of α over $\mathbb{Q}$. Thus we let $P(x) \in \mathbb{Z}[x]$ be a polynomial of degree d with $P(\alpha) = 0$.

Step II: The Polynomial Must Vanish at p/q

Suppose that p/q closely approximates α and that q is large. We want to show that $P(p/q) = 0$.

First we use the fact that $P(x)$ has degree d and has integer coefficients to deduce that

$$P\left(\frac{p}{q}\right) = \frac{N}{q^d} \qquad \text{for some integer } N \in \mathbb{Z}.$$

Next we use Taylor's theorem and the triangle inequality to give an upper bound for $P(p/q)$. We expand $P(x)$ around $x = \alpha$ as

$$P(x) = \sum_{i=1}^{d} \frac{1}{i!} \frac{d^i P}{dx^i}(\alpha)(x-\alpha)^i.$$

Note that there is no constant term, since $P(\alpha) = 0$, so if p/q satisfies the inequality $(*)$, then $P(p/q)$ will be small. Explicitly, we get the estimate

$$\begin{aligned}\left|\frac{N}{q^d}\right| = \left|P\left(\frac{p}{q}\right)\right| &\le \left|\frac{p}{q} - \alpha\right| \cdot \left(\sum_{i=1}^{d} \left|\frac{1}{i!}\frac{d^i P}{dx^i}(\alpha)\right| \cdot \left|\frac{p}{q} - \alpha\right|^{i-1}\right) \\ &\le \left|\frac{p}{q} - \alpha\right| \cdot d \cdot \max_{1 \le i \le d} \left|\frac{1}{i!}\frac{d^i P}{dx^i}(\alpha)\right| = B(\alpha)\left|\frac{p}{q} - \alpha\right| \le \frac{B(\alpha)}{q^{d+\varepsilon}}.\end{aligned}$$

Here $B(\alpha)$ is a positive constant that depends only on α.

It follows that $|N| \le B(\alpha)/q^{\varepsilon}$. But N is an integer, and there are no integers strictly between 0 and 1. This proves:

$$\text{If } \frac{p}{q} \text{ satisfies } (*) \text{ and } q > B(\alpha)^{1/\varepsilon}, \text{ then } P\left(\frac{p}{q}\right) = 0.$$

Step III: The Polynomial Does Not Vanish at p/q

In this third step we need to verify that $P(p/q)$ is not zero, which will contradict the conclusion of Step II. In our case, the nonvanishing of $P(p/q)$ is trivial, since $P(x)$ is the minimal polynomial of α over $\mathbb{Q}$, so $P(x)$ is

irreducible in $\mathbb{Q}[x]$ and thus certainly cannot have any rational roots. Do not be misled by the simplicity of this step. The theorems of Thue–Siegel–Roth require the use of polynomials of more than one variable, and then the nonvanishing step becomes the most difficult part of the proof.

Step IV: Completion of the Proof

We suppose that the inequality (*) has infinitely many solutions $p/q \in \mathbb{Q}$ and derive a contradiction. Under this hypothesis, we can choose a solution p_1/q_1 whose denominator satisfies $q_1 > B(\alpha)^{1/\varepsilon}$, where $B(\alpha)$ is the constant described in Step II. Next we let $P(x)$ be the polynomial constructed in Step I. From Step II we see that $P(p_1/q_1) = 0$, while Step III tells us that $P(p_1/q_1) \neq 0$. This contradiction tells us that (*) can have only finitely many solutions. □

Remark D.1.2.1. Liouville's estimate (D.1.2) says that an algebraic number cannot be too closely approximated by rational numbers. Liouville used this result to prove the existence of transcendental numbers. (See Exercise D.5.) However, the exponent $d+\varepsilon$ is too large for most applications to Diophantine equations. For example, in order to prove that the equation

$$x^d - 2y^d = m$$

has only finitely many solutions in integers $x, y \in \mathbb{Z}$, one needs Liouville's estimate with an exponent of $d-\varepsilon$. The improvements of Liouville's result given by Thue, Siegel, and Roth thus have profound Diophantine consequences.

Remark D.1.2.2. As we have stated it, Liouville's result (D.1.2) gives an exponent $d+\varepsilon$. The same argument actually gives an effective constant $C(\alpha) > 0$ such that

$$\left|\frac{p}{q} - \alpha\right| \geq \frac{C(\alpha)}{q^d} \qquad \text{for all } \frac{p}{q} \in \mathbb{Q}.$$

(See Exercise D.4.) We have elected to state Liouville's theorem with the weaker exponent $d+\varepsilon$ in order to emphasize its relationship to the other results in this chapter.

Remark D.1.2.3. Dirichlet's estimate (D.1.1) says that every irrational number α can be approximated by rationals to within $1/q^2$. It is natural to ask whether for algebraic numbers this is the best possible result. Thus if α is an irrational algebraic number, does there exist a constant $c(\alpha) > 0$ such that

$$\left|\frac{p}{q} - \alpha\right| \geq \frac{c(\alpha)}{q^2} \qquad \text{for all } \frac{p}{q} \in \mathbb{Q}?$$

Liouville's result (D.1.2) (see also remark D.1.2.2 above) says that this is true if α is quadratic, but if α is an algebraic number of degree strictly greater than 2, then it is conjectured to be false. However, there is not a single such number (e.g., $\sqrt[3]{2}$) for which it is known to be false. See Exercise D.18 for further information.

D.2. Roth's Theorem

Roth's theorem, as stated in the introduction to this chapter, asserts that every algebraic number α has approximation exponent 2. That is, for any $\varepsilon > 0$, there are only finitely many rational number p/q satisfying

$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{q^{2+\varepsilon}}.$$

More generally, one might allow the approximating values to be taken from a number field other than $\mathbb{Q}$, and one might replace the single archimedean absolute value with several absolute values. This leads to the following general formulation of Roth's theorem, which we will prove in this chapter.

Theorem D.2.1. (Roth's theorem) *Let K be a number field, let $S \subset M_K$ be a finite set of absolute values on K, and assume that each absolute value in S has been extended in some way to $\bar{K}$. Let $\alpha \in \bar{K}$ and $\varepsilon > 0$ be given. Then there are only finitely many $\beta \in K$ satisfying the inequality*

$$\prod_{v \in S} \min\{\|\beta - \alpha\|_v, 1\} \le \frac{1}{H_K(\beta)^{2+\varepsilon}}. \qquad (*)$$

The numerous details required for the proof of Roth's theorem will occupy us for the next several sections. To assist the reader, we begin with a brief overview of the proof, which proceeds by contradiction. For simplicity, we will assume that S contains a single absolute value v. So we suppose that there are infinitely many solutions $\beta \in K$ to the inequality $(*)$. We choose solutions $\beta_1, \beta_2, \dots, \beta_m$ to $(*)$ with the property that $H_K(\beta_1)$ is large, and each $H_K(\beta_{i+1})$ is much larger than $H_K(\beta_i)$. Next we construct a polynomial $P(X_1, \dots, X_m)$ with integer coefficients that vanishes to high order at the point $(\alpha, \alpha, \dots, \alpha)$. Using the Taylor expansion of P around $(\alpha, \dots, \alpha)$ and the fact that the β_i's are close to α, we see that $\|P(\beta_1, \dots, \beta_m)\|_v$ is so small that it must vanish, and the same is true of many of its derivatives. In other words, P must vanish to fairly high order at $(\beta_1, \dots, \beta_m)$. Finally we apply Roth's lemma, which says that P cannot vanish to high order at $(\beta_1, \dots, \beta_m)$, to obtain the desired contradiction. Roth's lemma, which is proven in Section D.6, is technically the most difficult part of the proof.

Before beginning the proof of Roth's theorem, we are going to make two simplifications. The first is very elementary and says that it suffices to prove Roth's theorem for algebraic integers. This will make some of our later calculations easier.

Reduction Lemma D.2.1.1. *If Roth's theorem (D.2.1) is true for all algebraic integers, then it is true for all algebraic numbers.*

PROOF. Let α be an algebraic number, and suppose that Roth's theorem is false for α. This means that there are infinitely many $\beta \in K$ satisfying the inequality $(*)$. The set S has only finitely many subsets, so possibly after replacing S by one of its subsets, we may assume that there are infinitely many $\beta \in K$ such that

$$\prod_{v\in S} \|\beta - \alpha\|_v \le \frac{1}{H_K(\beta)^{2+\varepsilon}}.$$

Choose an integer $D > 0$ such that $D\alpha$ is an algebraic integer, and let $\beta \in K$ be a solution to $(*)$ with $H_K(\beta) > H_K(D)^{1+6/\varepsilon}$. It is clear from the definition of the height that $H_K(D\beta) \le H_K(D)H_K(\beta)$. Further,

$$\prod_{v\in S} \|D\|_v \le \prod_{v\in S} \max\{\|D\|_v, 1\} = H_K(D).$$

Hence

$$\begin{aligned}\prod_{v\in S} \|D\beta - D\alpha\|_v &\le \frac{H_K(D)}{H_K(\beta)^{2+\varepsilon}} \\ &= \frac{H_K(D)}{H_K(\beta)^{2+\varepsilon/2}} \cdot \frac{1}{H_K(\beta)^{\varepsilon/2}} \\ &\le \frac{H_K(D)}{\big(H_K(D\beta)/H_K(D)\big)^{2+\varepsilon/2}} \cdot \frac{1}{\big(H_K(D)^{1+6/\varepsilon}\big)^{\varepsilon/2}} \\ &= \frac{1}{H_K(D\beta)^{2+\varepsilon/2}}.\end{aligned}$$

Thus $D\beta$ is a close approximation to $D\alpha$ in the sense that the inequality $(*)$ is true when $\alpha, \beta, \varepsilon$ are replaced by $D\alpha, D\beta, \varepsilon/2$. Hence the falsity of Roth's theorem for α implies its falsity for the algebraic integer $D\alpha$. This proves that if Roth's theorem is true for algebraic integers, then it is true for all algebraic numbers. □

The next theorem closely resembles Roth's theorem, but it replaces the condition that the product $\prod \|\beta - \alpha\|_v$ be small with the condition that each of the differences $\|\beta - \alpha\|_v$ be small. This idea of reduction to simultaneous approximation is due to Mahler [1], who was also the first one to study Diophantine approximation for p-adic absolute values.

Theorem D.2.2. *Let K be a number field, let $S \subset M_K$ be a finite set of absolute values on K with each absolute value extended in some way to $\bar{K}$. Let $\alpha \in \bar{K}$ and $\varepsilon > 0$ be given. Suppose that*

$$\xi : S \to [0,1] \quad \textit{is a function satisfying} \quad \sum_{v\in S} \xi_v = 1.$$

Then there are only finitely many $\beta \in K$ with the property that

$$\|\beta - \alpha\|_v \le \frac{1}{H_K(\beta)^{(2+\epsilon)\xi_v}} \quad \textit{for all } v \in S. \qquad (**)$$

It is fairly clear that Theorem D.2.1 implies Theorem D.2.2, but the converse is a bit trickier.

Reduction Lemma D.2.2.1. *Theorem D.2.1 is true if and only if Theorem D.2.2 is true.*

PROOF. Suppose first that Theorem D.2.1 is true. Let $\xi : S \to [0,1]$ be a function as described in Theorem D.2.2, and suppose that $\beta \in K$ satisfies (∗∗). Multiplying the estimate (∗∗) over $v \in S$ and using $\sum_v \xi_v = 1$ shows that β satisfies the inequality (∗), so Theorem D.2.1 tells us that there are only finitely many β's.

Next suppose that Theorem D.2.2 is true, and suppose that there are infinitely many numbers $\beta \in K$ satisfying (∗). We will derive a contradiction, which will prove the other implication. The idea is to construct several functions ξ of the sort described in Theorem D.2.2 so that each β satisfies (∗∗) for at least one of our ξ's.

Let $s = \#S$. We consider the collection of maps

$$\xi : S \to [0,1] \quad \text{of the form } \xi_v = \frac{a_v}{s} \text{ with } a_v \in \mathbb{Z},\ a_v \ge 0, \text{ and } \sum_{v \in S} a_v = s.$$

It is clear that there are only finitely many such maps. We will denote this collection of maps by $\mathcal{Z}$.

Now suppose that $\beta \in K$ satisfies (∗). We want to show that β satisfies (∗∗) for one of the maps in $\mathcal{Z}$. For each $v \in S$, define a real number $\lambda_v(\beta) \ge 0$ by the formula

$$\min\{\|\beta - \alpha\|_v, 1\} = \frac{1}{H_K(\beta)^{(2+\epsilon)\lambda_v(\beta)}}.$$

Multiplying over $v \in S$ and comparing with (∗), we see that $\sum_{v \in S} \lambda_v(\beta) \ge 1$, so

$$\sum_{v \in S} \lceil 2s\lambda_v(\beta) \rceil \ge \sum_{v \in S} \bigl(2s\lambda_v(\beta) - 1\bigr) = 2s \sum_{v \in S} \lambda_v(\beta) - s \ge s.$$

This implies that we can find integers $b_v(\beta)$ with the property that

$$0 \le b_v(\beta) \le 2s\lambda_v(\beta) \qquad \text{and} \qquad \sum_{v \in S} b_v(\beta) = s.$$

Then the function $\xi : S \to [0,1]$ defined by $\xi_v = b_v(\beta)/s$ is in the set $\mathcal{Z}$.

We have now proven that if $\beta \in K$ satisfies $(*)$, then β satisfies $(**)$ for at least one of the functions ξ in $\mathcal{Z}$. But by assumption, for any given ξ there are only finitely many β's satisfying $(**)$, and it is clear that $\mathcal{Z}$ contains only finitely many functions. Therefore $(*)$ has finitely many solutions. This completes the proof that Theorem D.2.2 implies Theorem D.2.1.

□

D.3. Preliminary Results

In this section we will prove a number of preliminary results that will be needed at various points during the proof of Roth's theorem. Other than (D.3.6), none of these results is particularly difficult to prove, but taken together they form a rather lengthy whole. So we urge the reader to study only the definitions and statements of results for now, and then proceed to the actual proof of Roth's theorem that begins in the next section. To facilitate this procedure, we have collected the main results at the beginning and relegated the proofs to the end of the section.

We will be working especially with polynomials having integer coefficients, since ultimately all proofs in Diophantine approximation, from Liouville's elementary result to Roth's theorem, depend on the fact that there are no integers strictly between 0 and 1. We will be concerned with both the size of the coefficients and the partial derivatives of these polynomials, which prompts us to set the following notation.

Definition. Let $P(X_1,\dots,X_m) \in \mathbb{R}[X_1,\dots,X_m]$ be a polynomial and let $(i_1,\dots,i_m)$ be an m-tuple of nonnegative integers. We define

$$|P| = \text{maximum absolute value of coefficients of } P,$$

$$\partial_{i_1\dots i_m}P = \frac{1}{i_1!i_2!\cdots i_m!}\frac{\partial^{i_1+\cdots+i_m}}{\partial X_1^{i_1}\cdots X_m^{i_m}}P.$$

The normalized partial derivative $\partial_{i_1\dots i_m}P$ is designed to cancel, as much as possible, the common factors that appear in the coefficients when we differentiate. This is the content of part (a) of the following elementary lemma, while part (b) gives a bound for the coefficients of the derivative.

Lemma D.3.1. *Let $P(X_1,\dots,X_m) \in \mathbb{Z}[X_1,\dots,X_m]$ be a polynomial with integer coefficients, and let $= (i_1,\dots,i_m)$ be an m-tuple of nonnegative integers.*

(a) $\partial_{i_1\dots i_m}P \in \mathbb{Z}[X_1,\dots,X_m]$.

(b) *If $\deg_{X_h}(P) \le r_h$ for each $1 \le h \le m$, then*

$$|\partial_{i_1\dots i_m}P| \le 2^{r_1+\cdots+r_m}|P|.$$

Remark D.3.1.1. Lemma D.3.1(b) gives one reason why we use $\partial_{i_1\dots i_m}$ instead of the more natural operator $D_{i_1\dots i_m} = \partial^{i_1+\cdots+i_m}/\partial X_1^{i_1}\cdots X_m^{i_m}$. Both operators map the polynomial ring $\mathbb{Z}[X_1,\dots,X_m]$ to itself. However, if we use $D_{i_1\dots i_m}$, then Lemma D.3.1(b) would be replaced by the weaker estimate

$$|D_{i_1\dots i_m}P| \le (r_1!)(r_2!)\cdots(r_m!)|P|.$$

Now $r!$ is approximately $(r/e)^r$ by Sterling's formula. The reader may check that the improvement from $r!$ to 2^r is vital in the estimates used in the proof of Roth's theorem.

There are other reasons to prefer the ∂ operators. Taylor's formula is simpler,

$$P(X_1,\dots,X_m) = \sum_{i_1=1}^{r_1}\cdots\sum_{i_m=1}^{r_m}\partial_{i_1\dots i_m}P(a_1,\dots,a_m)(X_1-a_1)^{i_1}\cdots(X_m-a_m)^{i_m}.$$

Leibniz's formula for the derivative of a product is also nicer,

$$\partial_n\left(P_1(X)P_2(X)\cdots P_s(X)\right) = \sum_{j_1+\cdots+j_s=n}\partial_{j_1}P_1(X)\cdots\partial_{j_s}P_s(X),$$

as is Cauchy's bound that now reads $|\partial_n P(a)| \le Mr^{-n}$, where $M = \max_{|z-a|\le r}|P(z)|$. We also point out that (even in characteristic p) we have

$$(X-a)^r \text{ divides } P(X) \iff \partial_i P(a) = 0 \text{ for all } i < r.$$

The only real advantage of the D operators is that they satisfy the relation $D_{i_1\dots i_m}D_{j_1\dots j_m} = D_{i_1+j_1\dots i_m+j_m}$, while the composition of ∂ operators is somewhat more complicated.

The key to proving Roth's theorem is estimating the order to which a polynomial in many variables vanishes at certain points. It is thus important to have a means of measuring this order of vanishing. Of course, by allowing the degree of a polynomial to be high, we can force it to vanish to high order, so it makes sense to look at some sort of weighted order of vanishing. This idea serves to motivate the following definition.

Definition. Let k be any field, let $P(X_1,\dots,X_m) \in k[X_1,\dots,X_m]$ be a polynomial, let $(\alpha_1,\dots,\alpha_m) \in k^m$ be a point, and let $r_1,\dots,r_m$ be positive integers. The *index of P with respect to* $(\alpha_1,\dots,\alpha_m;r_1,\dots,r_m)$, denoted by $\operatorname{Ind} P$, is the smallest value of

$$\frac{i_1}{r_1}+\frac{i_2}{r_2}+\cdots+\frac{i_m}{r_m}$$

such that

$$\partial_{i_1\dots i_m}P(\alpha_1,\dots,\alpha_m) \ne 0.$$

(If P is the zero polynomial, we set the index equal to ∞.)

We make two observations. First, $\operatorname{Ind} P \geq 0$, with equality if and only if $P(\alpha_1, \ldots, \alpha_m) \neq 0$. Second, for any nonzero polynomial of multidegree less than or equal to $(r_1, \ldots, r_m)$, we have $\operatorname{Ind} P \leq m$, with equality if and only if

$$P(X_1, \ldots, X_m) = c(X_1 - \alpha_1)^{r_1} \cdots (X_m - \alpha_m)^{r_m}.$$

Remark. There is no connection a priori between the weights r_i with respect to which the index is computed and the degrees of P, but in practice the index is always applied to polynomials P with $\deg_{X_i} P \leq r_i$.

The following lemma gives some elementary properties satisfied by the index. Notice that (b) and (c) say that the index is a discrete valuation on the polynomial ring $k[X_1, \ldots, X_m]$, generalizing the usual "order of vanishing" valuation on $k[X]$.

Lemma D.3.2. *Let $P, P' \in k[X_1, \ldots, X_m]$ be polynomials, and fix integers $r_1, \ldots, r_m$ and a point $(\alpha_1, \ldots, \alpha_m) \in k^m$. The index with respect to $(\alpha_1, \ldots, \alpha_m; r_1, \ldots, r_m)$ has the following properties:*

(a) $\operatorname{Ind}(\partial_{i_1 \ldots i_m} P) \geq \operatorname{Ind} P - \left(\dfrac{i_1}{r_1} + \cdots + \dfrac{i_m}{r_m}\right)$.

(b) $\operatorname{Ind}(P + P') \geq \min\{\operatorname{Ind} P, \operatorname{Ind} P'\}$.

(c) $\operatorname{Ind}(PP') = \operatorname{Ind} P + \operatorname{Ind} P'$.

The following important lemma is in essence nothing more than the assertion that there are no integers strictly between 0 and 1. (To see why, consider what it says for $K = \mathbb{Q}$, $S = M_{\mathbb{Q}}^{\infty}$, and $\alpha \in \mathbb{Q}$ a rational number with $|\alpha| \leq 1$.)

Lemma D.3.3. (Liouville's inequality) *Let $K/\mathbb{Q}$ be a number field, let $\alpha \in K^*$ be a nonzero element of K, and let $S \subset M_K$ be any set of absolute values on K. Then*

$$\prod_{v \in S} \min\{\|\alpha\|_v, 1\} \geq \frac{1}{H_K(\alpha)}.$$

If α is an algebraic integer of degree d over $\mathbb{Q}$, then every element in the ring $\mathbb{Z}[\alpha]$ can be uniquely written as a $\mathbb{Z}$-linear combination of the basis elements $1, \alpha, \alpha^2, \ldots, \alpha^{d-1}$. The next lemma tells us how large the coefficients become when we write a power α^ℓ as such a linear combination.

Lemma D.3.4. *Let α be an algebraic integer of degree d over $\mathbb{Q}$, and let*

$$Q(X) = X^d + a_1 X^{d-1} + \cdots + a_{d-1} X + a_d \in \mathbb{Z}[X]$$

be the minimal polynomial of α over $\mathbb{Q}$. Then for every $\ell \geq 0$ we can write

$$\alpha^\ell = a_1^{(\ell)} \alpha^{d-1} + a_2^{(\ell)} \alpha^{d-2} + \cdots + a_{d-1}^{(\ell)} \alpha + a_d^{(\ell)}$$

with integers

$$a_i^{(\ell)} \in \mathbb{Z} \qquad \textit{satisfying} \qquad \left|a_i^{(\ell)}\right| \le \left(|Q|+1\right)^{\ell}.$$

In order to create a polynomial P whose index is large at some point, we will have to choose the coefficients of P such that many of its derivatives are equal to 0 at that point. The next two lemmas, which are purely combinatorial in nature, will help us determine the number of conditions that the coefficients of P must satisfy. The first is very elementary; it counts the number of monomials of degree r in m variables and has been implicitly used in Parts A and B.

Lemma D.3.5. *Let $m \ge 1$ and $r \ge 0$ be given integers. Then there are exactly*

$$\binom{r+m-1}{r}$$

m-tuples of integers $(i_1, \dots, i_m)$ satisfying the conditions

$$i_h \ge 0 \textit{ for all } 1 \le h \le m, \qquad \textit{and} \qquad i_1 + i_2 + \cdots + i_m = r.$$

The next combinatorial lemma, which is somewhat more difficult to prove, admits a probabilistic interpretation. If we "randomly" choose an m-tuple $(i_1, \dots, i_m)$ with $0 \le i_h \le r_h$, then we would expect each i_h/r_h to be approximately equal to $\frac{1}{2}$, and so we would expect $\sum i_h/r_h$ to be approximately $m/2$. Lemma D.3.6 quantifies this intuition by saying that the probability of a random m-tuple $(i_1, \dots, i_m)$ satisfying

$$\frac{1}{m}\sum_{h=1}^{m} \frac{i_h}{r_h} < \frac{1}{2} - \varepsilon$$

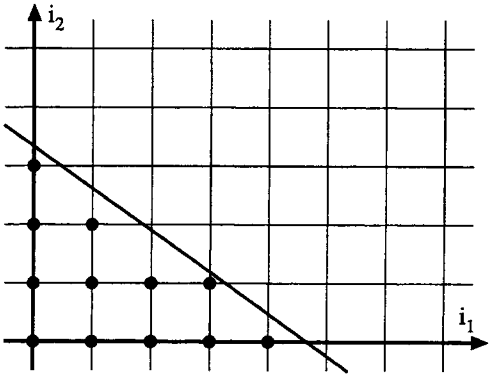
is at most $e^{-\varepsilon^2 m/4}$. For large values of m, this probability will be quite small. This type of estimate is a version of Chebyshev's inequality in probability theory. See Figure D.1 for an illustration with $m = 2$. It should also be pointed out that the average value $\frac{1}{2}$ is quite important, since it is this value that "explains" the 2 in Roth's theorem!

Lemma D.3.6. *Let $r_1, \dots, r_m$ be positive integers and fix an $0 < \varepsilon < 1$. Then there are at most*

$$(r_1+1)\cdots(r_m+1)\cdot e^{-\varepsilon^2 m/4}$$

m-tuples of integers $(i_1, \dots, i_m)$ in the range

$$0 \le i_1 \le r_1, \quad 0 \le i_2 \le r_2, \quad \dots, \quad 0 \le i_m \le r_m$$



Lattice points satisfying $\frac{1}{2}\left(\frac{i_1}{r_1}+\frac{i_2}{r_2}\right) \le \frac{1}{2} - \varepsilon$

Figure D.1

that satisfy the condition

$$\frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m} \le \frac{m}{2} - \varepsilon m.$$

This concludes the description of the preliminary results needed for the proof of Roth's theorem. The remainder of this section is devoted to the proofs of these results.

PROOF (of Lemma D.3.1). The elementary but crucial observation is

$$\frac{1}{i!}\frac{d^i X^j}{dX^i} = \frac{j(j-1)\cdots(j-i+1)}{i!}X^{i-j} = \binom{j}{i}X^{i-j},$$

where the combinatorial symbol

$$\binom{j}{i} = \frac{j!}{i!(j-i)!}$$

is an integer. If $j < i$, we define $\binom{j}{i}$ to be zero as usual.

Write the polynomial P as

$$P(X_1,\ldots,X_m) = \sum_{j_1=0}^{r_1}\cdots\sum_{j_m=0}^{r_m} p_{j_1,\ldots,j_m} X_1^{j_1}\cdots X_m^{j_m}.$$

Differentiating P, we obtain

$$\begin{aligned}\partial_{i_1\ldots i_m}P &= \sum_{j_1=0}^{r_1}\cdots\sum_{j_m=0}^{r_m} p_{j_1,\ldots,j_m}\left(\frac{1}{i_1!}\frac{\partial^{i_1}X_1^{j_1}}{\partial X_1^{i_1}}\right)\cdots\left(\frac{1}{i_m!}\frac{\partial^{i_m}X_m^{j_m}}{\partial X_m^{i_m}}\right)\\ &= \sum_{j_1=0}^{r_1}\cdots\sum_{j_m=0}^{r_m} p_{j_1,\ldots,j_m}\binom{j_1}{i_1}\cdots\binom{j_m}{i_m}X_1^{j_1-i_1}\cdots X_m^{j_m-i_m}.\end{aligned}$$

The combinatorial symbols are integers, so this proves part (a).

To prove (b), we use the binomial formula for $(1+1)^j$ to estimate

$$\binom{j}{i} \le \sum_{k=0}^{j} \binom{j}{k} = (1+1)^j = 2^j.$$

Hence taking the maximum over all m-tuples $(j_1, \dots, j_m)$ of integers satisfying

$$0 \le j_1 \le r_1, \quad \dots \quad 0 \le j_m \le r_m,$$

we find that

$$\begin{aligned} |\partial_{i_1 \dots i_m} P| &= \max \left| p_{j_1,\dots,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} \right| \\ &\le \max |p_{j_1,\dots,j_m}| \cdot \max 2^{j_1+\cdots+j_m} = |P| \cdot 2^{r_1+\cdots+r_m}. \end{aligned}$$

This completes the proof of (b). □

PROOF (of Lemma D.3.2). To ease notation, we will write

$$\alpha = (\alpha_1, \dots, \alpha_m).$$

(a) Let $Q = \partial_{i_1 \dots i_m} P$. Using the definition of the index, we can choose an m-tuple $(j_1, \dots, j_m)$ such that $\partial_{j_1 \dots j_m} Q(\alpha) \ne 0$ and such that the index of Q at $(\alpha_1, \dots, \alpha_m; r_1, \dots, r_m)$ is equal to $j_1/r_1 + \cdots + j_m/r_m$. Then

$$\begin{aligned} \partial_{j_1 \dots j_m} Q(\alpha) \ne 0 &\Longrightarrow \partial_{i_1+j_1,\dots,i_m+j_m} P(\alpha) \ne 0 \\ &\Longrightarrow \frac{i_1+j_1}{r_1} + \cdots + \frac{i_m+j_m}{r_m} \ge \operatorname{Ind} P \\ &\Longrightarrow \operatorname{Ind} Q = \frac{j_1}{r_1} + \cdots + \frac{j_m}{r_m} \ge \operatorname{Ind} P - \frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m}. \end{aligned}$$

(b) Choose an m-tuple $(j_1, \dots, j_m)$ such that $\partial_{j_1 \dots j_m}(P+P')(\alpha) \ne 0$ and the index of $P+P'$ at $(\alpha_1, \dots, \alpha_m; r_1, \dots, r_m)$ is equal to $j_1/r_1 + \cdots + j_m/r_m$. Then either $\partial_{j_1 \dots j_m} P(\alpha) \ne 0$ or $\partial_{j_1 \dots j_m} P'(\alpha) \ne 0$ (or both), which implies that $j_1/r_1 + \cdots + j_m/r_m$ is greater than or equal to at least one of $\operatorname{Ind} P$ or $\operatorname{Ind} P'$. Hence

$$\operatorname{Ind}(P+P') = \frac{j_1}{r_1} + \cdots + \frac{j_m}{r_m} \ge \min\{\operatorname{Ind} P, \operatorname{Ind} P'\}.$$

(c) Using the product rule, we can write

$$\partial_{j_1,\dots,j_m}(PP') = \sum_{i_1+i_1'=j_1} \cdots \sum_{i_m+i_m'=j_m} C_{i_1,\dots,i_m'} (\partial_{i_1 \dots i_m} P)(\partial_{i_1' \dots i_m'} P')$$

for certain positive integers $C_{i_1,\dots,i'_m}$. (In fact, all of the C's are equal to 1; see Exercise D.7 or the remarks on derivations.)

Choose an m-tuple $(j_1,\dots,j_m)$ such that $\partial_{j_1\dots j_m}(PP')(\alpha) \neq 0$ and such that

$$\operatorname{Ind}(PP') = \frac{j_1}{r_1} + \cdots + \frac{j_m}{r_m} \quad \text{at } (\alpha_1,\dots,\alpha_m; r_1,\dots,r_m).$$

Then there exist m-tuples $(i_1,\dots,i_m)$ and $(i'_1,\dots,i'_m)$ with $\partial_{i_1\dots i_m}P(\alpha) \neq 0$ and $\partial_{i'_1\dots i'_m}P'(\alpha) \neq 0$. Hence

$$\operatorname{Ind} P \le \sum_{h=1}^{m} \frac{i_h}{r_h} \qquad \text{and} \qquad \operatorname{Ind} P' \le \sum_{h=1}^{m} \frac{i'_h}{r'_h},$$

and adding these inequalities gives $\operatorname{Ind} P + \operatorname{Ind} P' \le \operatorname{Ind}(PP')$.

To get the inequality in the other direction, we look at the set of m-tuples $(i_1,\dots,i_m)$ satisfying

$$\partial_{i_1\dots i_m}P(\alpha) \neq 0 \qquad \text{and} \qquad \operatorname{Ind} P = \sum_{h=1}^{m} \frac{i_h}{r_h}.$$

We order these m-tuples lexicographically and choose the smallest one, call it $(\bar{\imath}_1,\dots,\bar{\imath}_m)$. This means that if $(i_1,\dots,i_m)$ is another such m-tuple, then there exists a $k \ge 1$ such that

$$i_k = \bar{\imath}_k \text{ for all } 1 \le h < k, \text{ and } i_k > \bar{\imath}_k.$$

We similarly choose an m-tuple $(\bar{\imath}'_1,\dots,\bar{\imath}'_m)$ for P', and we set

$$(\bar{\jmath}_1,\dots,\bar{\jmath}_m) = (\bar{\imath}_1 + \bar{\imath}'_1,\dots,\bar{\imath}_m + \bar{\imath}'_m).$$

Then

$$\partial_{\bar{\jmath}_1\dots\bar{\jmath}_m}(PP')(\alpha) = C_{\bar{\imath}_1,\dots,\bar{\imath}'_m}\partial_{\bar{\imath}_1\dots\bar{\imath}_m}P(\alpha)\cdot\partial_{\bar{\imath}'_1\dots\bar{\imath}'_m}P'(\alpha) \neq 0,$$

since all of the other terms will be zero. Therefore,

$$\operatorname{Ind}(PP') \le \sum_{h=1}^{m} \frac{\bar{\jmath}_h}{r_h} = \sum_{h=1}^{m} \frac{\bar{\imath}_h + \bar{\imath}'_h}{r_h} = \operatorname{Ind} P + \operatorname{Ind} P'.$$

This is the other inequality, which completes the proof that $\operatorname{Ind}(PP') = \operatorname{Ind} P + \operatorname{Ind} P'$. ☐

Proof (of Lemma D.3.3). The product formula (B.1.2) says that

$$\prod_{v\in M_K} \|\alpha\|_v = 1, \quad \text{which implies that} \quad 1 = \prod_{v\in M_K} \frac{1}{\|\alpha\|_v}.$$

(This is where we use the fact that α is nonzero.) Using this, we compute

$$\begin{aligned} H_K(\alpha) &= \prod_{v\in M_K} \max\{\|\alpha\|_v, 1\} = \prod_{v\in M_K} \|\alpha\|_v \cdot \max\left\{1, \frac{1}{\|\alpha\|_v}\right\} \\ &= \prod_{v\in M_K} \max\left\{1, \frac{1}{\|\alpha\|_v}\right\} = \prod_{v\in M_K} \frac{1}{\min\{1, \|\alpha\|_v\}} \ge \prod_{v\in S} \frac{1}{\min\{1, \|\alpha\|_v\}}. \end{aligned}$$

Taking reciprocals gives the desired result. □

PROOF (of Lemma D.3.4). The proof is by induction on ℓ. The assertion is clearly true for $0 \le \ell \le d-1$, since for ℓ in this range we can take all of the $a_i^{(\ell)}$'s to be either 0 or 1.

Assume now that the lemma is true for α^ℓ. We compute

$$\begin{aligned} \alpha^{\ell+1} &= \alpha\cdot\alpha^\ell = \alpha \sum_{i=1}^{d} a_i^{(\ell)} \alpha^{d-i} \\ &= a_1^{(\ell)}\alpha^d + \sum_{i=2}^{d} a_i^{(\ell)}\alpha^{d+1-i} \\ &= a_1^{(\ell)} \sum_{i=1}^{d} -a_i\alpha^{d-i} + \sum_{i=2}^{d} a_i^{(\ell)}\alpha^{d+1-i} \quad \text{using } Q(\alpha)=0 \\ &= \sum_{i=1}^{d} \left(-a_1^{(\ell)}a_i + a_{i+1}^{(\ell)}\right)\alpha^{d-i}, \quad \text{where we set } a_{d+1}^{(\ell)} = 0. \end{aligned}$$

It follows that

$$a_i^{(\ell+1)} = -a_1^{(\ell)}a_i + a_{i+1}^{(\ell)},$$

and so

$$\begin{aligned} \left|a_i^{(\ell+1)}\right| &\le \left|a_1^{(\ell)}a_i\right| + \left|a_{i+1}^{(\ell)}\right| \le \max\left\{\left|a_1^{(\ell)}\right|, \left|a_{i+1}^{(\ell)}\right|\right\}\cdot\left(|a_i|+1\right) \\ &\le \left(|Q|+1\right)^\ell \cdot \left(|Q|+1\right) \quad \text{by the induction hypothesis} \\ &= \left(|Q|+1\right)^{\ell+1}. \end{aligned}$$

□

PROOF (of Lemma D.3.5). This can be proven by a straightforward induction, but we will give a more illuminating counting argument. For any given m-tuple $(i_1, \dots, i_m)$, we replace each integer i_h by i_h dots, making sure to leave all of the commas in place. For example, we would represent the m-tuple $(3,2,0,4,0,0,3,0)$ (with $m=8$ and $r=12$) as follows:

$$(\bullet\bullet\bullet, \bullet\bullet, , \bullet\bullet\bullet\bullet, , , \bullet\bullet\bullet,).$$

The key observation is that the total number of "dots" and "commas" is $r+m-1=19$. In other words, to form an m-tuple with $i_1+\cdots+i_m=r$, we should start with a row of $r+m-1$ "commas" and change r of them into dots. Each choice of r commas to change will give us a unique m-tuple with the desired properties. Hence the total number of m-tuples is the number of ways of choosing r elements from an ordered set of $r+m-1$ objects. □

Proof (of Lemma D.3.6). Let $I(m,\varepsilon)$ be the set of m-tuples that we are trying to count,

$$I(m,\varepsilon)=\Big\{(i_1,\ldots,i_m)\in\mathbb{Z}^m\ :\ 0\le i_h\le r_h \text{ for all } 1\le h\le m \text{ and } \sum_{h=1}^{m}\frac{i_h}{r_h}\le\frac{m}{2}-\varepsilon m\Big\}.$$

Then

$$\begin{aligned}
\#I(m,\varepsilon) &= \sum_{(i_1,\ldots,i_m)\in I(m,\varepsilon)} 1\\
&\le \sum_{(i_1,\ldots,i_m)\in I(m,\varepsilon)} \exp\left[\frac{\varepsilon}{2}\left(\frac{m}{2}-\varepsilon m-\frac{i_1}{r_1}-\cdots-\frac{i_m}{r_m}\right)\right] && \text{since } e^t\ge 1 \text{ for all } t\ge 0\\
&\le \sum_{i_1=0}^{r_1}\cdots\sum_{i_m=0}^{r_m}\exp\left[\frac{\varepsilon}{2}\left(\frac{m}{2}-\varepsilon m-\frac{i_1}{r_1}-\cdots-\frac{i_m}{r_m}\right)\right] && \text{since includes extra positive terms}\\
&= \exp\left(-\frac{\varepsilon^2 m}{2}\right)\sum_{i_1=0}^{r_1}\cdots\sum_{i_m=0}^{r_m}\exp\left[\frac{\varepsilon}{2}\left(\frac{m}{2}-\frac{i_1}{r_1}-\cdots-\frac{i_m}{r_m}\right)\right]\\
&= \exp\left(-\frac{\varepsilon^2 m}{2}\right)\prod_{h=1}^{m}\left(\sum_{i=0}^{r_h}\exp\left(\frac{\varepsilon}{2}\left(\frac{1}{2}-\frac{i}{r_h}\right)\right)\right).
\end{aligned}$$

We use the inequality

$$e^t\le 1+t+t^2,\quad \text{valid for all } |t|\le 1,$$

to estimate one of the inner sums as

$$\begin{aligned}
\sum_{i=0}^{r}\exp\left(\frac{\varepsilon}{2}\left(\frac{1}{2}-\frac{i}{r}\right)\right) &\le \sum_{i=0}^{r}\left\{1+\frac{\varepsilon}{2}\left(\frac{1}{2}-\frac{i}{r}\right)+\frac{\varepsilon^2}{4}\left(\frac{1}{2}-\frac{i}{r}\right)^2\right\}\\
&= \sum_{i=0}^{r}\left\{\left(1+\frac{\varepsilon}{4}+\frac{\varepsilon^2}{16}\right)-\left(\frac{\varepsilon}{2}+\frac{\varepsilon^2}{4}\right)\frac{i}{r}+\frac{\varepsilon^2}{4}\frac{i^2}{r^2}\right\}\\
&= (r+1)\left(1+\frac{\varepsilon^2}{48}+\frac{\varepsilon^2}{12r}\right)\\
&\le (r+1)\left(1+\frac{\varepsilon^2}{4}\right) \quad \text{using } r\ge 1.
\end{aligned}$$

(Although it may look complicated, the main point of this calculation is to get an upper bound of the form $1 + c\varepsilon^2$ with $c < \frac{1}{2}$. In particular, it would not suffice to get a bound of the form $1 + c\varepsilon$.)

Substituting this estimate in above, we find that

$$\begin{aligned}\#I(m,\varepsilon) &\le \exp\left(-\frac{\varepsilon^2 m}{2}\right)\prod_{h=1}^{m}\left((r_h+1)\left(1+\frac{\varepsilon^2}{4}\right)\right)\\ &\le \exp\left(-\frac{\varepsilon^2 m}{2}\right)\prod_{h=1}^{m}\left((r_h+1)\exp\left(\frac{\varepsilon^2}{4}\right)\right) \qquad \text{using } 1+t \le e^t\\ &= (r_1+1)\cdots(r_m+1)\exp\left(-\frac{\varepsilon^2 m}{4}\right).\end{aligned}$$

□

D.4. Construction of the Auxiliary Polynomial

In this section we are going to construct a polynomial $P(X_1, \ldots, X_m)$ with reasonably small integer coefficients and the property that P vanishes to high order at $(\alpha, \alpha, \ldots, \alpha)$. The way that we will build P is by solving a system of linear equations with integer coefficients. Results that describe integer solutions to systems of linear equations are often named after Siegel, because he was the first to formalize this procedure.

Lemma D.4.1. (Siegel's lemma, first version) *Let $N > M$ be positive integers, and let*

$$\begin{array}{ccccccc} a_{11}T_1 & + & \cdots & + & a_{1N}T_N & = & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ a_{M1}T_1 & + & \cdots & + & a_{MN}T_N & = & 0 \end{array}$$

be a system of linear equations with integer coefficients not all zero. Then there is a solution $(t_1, \ldots, t_N)$ to this system of equations with $t_1, \ldots, t_N$ integers, not all zero, and satisfying

$$\max_{1\le i\le N} |t_i| \le \left(N \max_{\substack{1\le i\le M\\ 1\le j\le N}} |a_{ij}|\right)^{\frac{M}{N-M}}.$$

Although the conclusion of this lemma looks a bit messy, it is really saying something quite simple. The system of homogeneous linear equations has more variables than equations, so we know it has nontrivial solutions. Since the coefficients are integers, there will be rational solutions; and by clearing the denominators of the rational solutions, we can find

integer solutions. So it is obvious that there are nonzero integer solutions. The last part of the lemma then says that we can find some solution that is not too large; precisely, we can find a solution whose size is bounded in terms of the number of equations M, the number of variables N, and the size of the coefficients a_{ij}. This, too, is not surprising; so the real content of the lemma is the precise form of the bound. We remark for future reference that the bound has, up to a small negligible term, the shape

$$\log\bigl(\max|t_i|\bigr) \le (\text{log height of the equations}) \cdot \left(\frac{\text{number of equations}}{\text{dimension of solutions}}\right).$$

PROOF (of Siegel's lemma). For any vector $\mathbf{t} = (t_1, \dots, t_N) \in \mathbb{R}^N$, we let

$$|\mathbf{t}| = \max_{1 \le i \le N} |t_i|$$

be the largest of the absolute values of its coordinates. Similarly, we will let A be the matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{MN} \end{pmatrix} \qquad \text{and will write} \qquad |A| = \max_{\substack{1 \le i \le M \\ 1 \le j \le N}} |a_{ij}|.$$

So the statement of Siegel's lemma is that the equation $A\mathbf{t} = \mathbf{0}$ has a solution $\mathbf{t} \in \mathbb{Z}^N$, $\mathbf{t} \ne \mathbf{0}$, satisfying

$$|\mathbf{t}| < \bigl(N|A|\bigr)^{M/(N-M)}.$$

The idea of the proof is very simple. The number of integer vectors $\mathbf{t}$ in $\mathbb{Z}^N$ of norm less than B is roughly B^N, while the norm of $A\mathbf{t}$ is less than $N|A||\mathbf{t}|$; hence the image vectors vary (roughly) among at most $(|A|B)^M$ values. When B^N is larger than $(|A|B)^M$, the pigeonhole principle will provide two integer vectors $\mathbf{t}_1 \ne \mathbf{t}_2$ of norm less than B with $A\mathbf{t}_1 = A\mathbf{t}_2$. Then the difference $\mathbf{t} = \mathbf{t}_1 - \mathbf{t}_2$ provides a solution to our linear system having norm less than $2B$. In order to get improved constants, we refine the previous argument as follows.

For any real number a we set

$$a^+ = \max(a, 0) \qquad \text{and} \qquad a^- = \max(-a, 0),$$

so that $a = a^+ - a^-$ and $|a| = a^+ + a^-$. We also define linear forms $L_j(\mathbf{t}) = \sum_{i=1}^N a_{ji} t_i$ and set

$$L_j^+ = \sum_{i=1}^N a_{ji}^+, \qquad L_j^- = \sum_{i=1}^N a_{ji}^-, \qquad \text{and} \qquad L_j = L_j^+ + L_j^- = \sum_{i=1}^N |a_{ji}|.$$

Assuming $0 \le t_i \le B$, we deduce that the jth coordinate $L_j(\mathbf{t})$ of $A\mathbf{t}$ lies in the interval

$$-L_j^- B \le L_j(\mathbf{t}) \le L_j^+ B.$$

Taking B to be an integer, the number of integer vectors in the box $\prod_{j=1}^{M}[-L_j^- B, L_j^+ B]$ is equal to

$$\prod_{j=1}^{M}(L_j^- B + L_j^+ B + 1) = \prod_{j=1}^{M}(L_j B + 1),$$

while the number of integer vectors $\mathbf{t}$ with $0 \le t_i \le B$ is $(B+1)^N$. Hence if we choose B to satisfy

$$(B+1)^N > \prod_{j=1}^{M}(L_j B + 1), \qquad (*)$$

then the pigeonhole principle provides us with distinct integer vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ such that $L_j(\mathbf{t}_1) = L_j(\mathbf{t}_2)$ for all $1 \le j \le N$. Then the vector $\mathbf{t} = \mathbf{t}_1 - \mathbf{t}_2$ is an integer solution of our linear system satisfying $|\mathbf{t}| \le B$.

To complete the proof of Siegel's lemma, it remains only to verify that the value

$$B := \left\lfloor \left(N \max_{i,j} |a_{ij}|\right)^{M/(N-M)} \right\rfloor$$

satisfies condition $(*)$. For this value of B we have

$$B + 1 > (N \max |a_{ij}|)^{M/(N-M)},$$

and hence

$$(B+1)^N = (B+1)^M (B+1)^{N-M} > (B+1)^M (N \max |a_{ij}|)^M.$$

Now we observe that

$$L_j \le N \max_i |a_{ij}| \quad \text{and} \quad 1 \le N \max_{i,j} |a_{ij}|,$$

and hence

$$\prod_{j=1}^{M}(L_j B + 1) \le \left((B+1) N \max_{i,j} |a_{ij}|\right)^M \le (B+1)^N.$$

□

We will now apply the same sort of pigeonhole principle argument to solve linear equations with algebraic coefficients. If the coefficients lie in a number field of degree d, and if we have M equations in N unknowns, then choosing a basis for the number fields allows us to translate the problem into dM equations with coefficients in $\mathbb{Q}$. Thus the relevant linear algebra condition is now $dM < N$. The following proof is taken from a paper of Anderson and Masser [1]; for a slightly sharper estimate, see Exercise D.9 and the references listed there.

Lemma D.4.2. (Siegel's lemma, second form) *Let k be a number field with $d = [k : \mathbb{Q}]$, let $a_{ij} \in k$ be elements not all zero, and let $A := H(\dots, a_{ij}, \dots)$ be the height of the vector formed by the a_{ij}'s. Assume that $dM < N$. Then there exists a nonzero vector $x \in \mathbb{Z}^N$ such that*

$$\sum_{i=1}^{N} a_{ij}x_i = 0 \quad \text{for all } 1 \le j \le M, \text{ and} \qquad \max_{1\le i\le N} |x_i| \le (NA)^{dM/(N-dM)}.$$

We begin by computing how many algebraic numbers are contained in various boxes.

Claim. *Let k be a number field of degree d, fix an element $\alpha_v \in k$ for each $v \in M_k$, and let $c = \{c_v\}_v$ be a multiplicative M_k-constant. That is, $c_v \ge 1$ for all $v \in M_k$, and $c_v = 1$ for all but finitely many $v \in M_k$ (cf. the additive M_k-constants defined in Section B.8.) Set $C := \prod_v c_v$. Then*

$$\operatorname{card}\bigl\{x \in k \;\bigm|\; |x - \alpha_v|_v \le c_v \text{ for all } v \in M_k\bigr\} \le \Bigl(2C^{1/d} + 1\Bigr)^d.$$

Proof (of Claim). Call $\mathcal{T}$ the set whose cardinality we are trying to bound. Set

$$E = \prod_{v\in M_k^\infty} k_v = \mathbb{R}^{r_1} \times \mathbb{C}^{r_2},$$

and for $\alpha \in k$ and $\varepsilon > 0$, consider the box

$$B(\alpha, \varepsilon) = \bigl\{x \in E \;\bigm|\; |x_v - \sigma_v(\alpha)| < \varepsilon c_v \text{ for all } v \in M_k^\infty\bigr\}.$$

We first observe that if $\alpha, \beta \in \mathcal{T}$ and if we take $\varepsilon = \frac{1}{2}C^{-1/d}$, then the intersection $B(\alpha,\varepsilon) \cap B(\beta,\varepsilon)$ is empty. To verify this, suppose that x sits in both boxes. If v is archimedean, then

$$|\alpha - \beta|_v = |\sigma_v(\alpha) - \sigma_v(\beta)| \le |x - \sigma_v(\alpha)| + |x - \sigma_v(\beta)| < 2\varepsilon c_v;$$

and if v is nonarchimedean, then

$$|\alpha - \beta|_v \le \max\bigl\{|\alpha - \alpha_v|, |\beta - \alpha_v|\bigr\} \le c_v.$$

If follows that $\prod |\alpha - \beta|_v < (2\varepsilon)^d C = 1$, and then the product formula tells us that $\alpha = \beta$.

Now the disjointedness of the $B(\alpha, \varepsilon)$'s for $\alpha \in \mathcal{T}$ implies that

$$\operatorname{Vol}\Bigl(\bigcup_{\alpha\in\mathcal{T}} B(\alpha,\varepsilon)\Bigr) = \operatorname{card}(\mathcal{T}) \operatorname{Vol}\bigl(B(0,\varepsilon)\bigr) = \operatorname{card}(\mathcal{T})\varepsilon^d \operatorname{Vol}\bigl(B(0,1)\bigr).$$

Next, if $x \in B(\alpha, \varepsilon)$ with $\alpha \in \mathcal{T}$, then

$$|x_v - \sigma_v(\alpha_v)| \leq |x_v - \sigma_v(\alpha)| + |\alpha - \alpha_v|_v \leq (1+\varepsilon)c_v.$$

These inequalities define a box with volume equal to $(1+\varepsilon)^d \operatorname{Vol}(B(0,1))$; hence

$$\operatorname{card}(\mathcal{T}) \leq \left(\frac{1+\varepsilon}{\varepsilon}\right)^d = \left(2C^{1/d}+1\right)^d.$$

□

PROOF (of Lemma D.4.2). To ease notation, put $\delta = dM/(N-dM)$ and $X = \left[(NA)^\delta\right]$. We apply the previous lemma with

$$\alpha_v = \begin{cases} L_j(X/2,\dots,X/2) & \text{if } v \text{ is archimedean,} \\ 0 & \text{otherwise;} \end{cases}$$
$$c_v = \begin{cases} NX \max|a_{ij}|_v/2 & \text{if } v \text{ is archimedean,} \\ \max|a_{ij}|_v & \text{otherwise.} \end{cases}$$

We then compute the associated "constant"

$$C = (NX/2)^d \prod_v \max|a_{ij}|_v \leq (NXA/2)^d.$$

We conclude that the linear form $L_j(x_1,\dots,x_N)$ takes at most $(1+NXA)^d$ values, and hence that L takes at most $(1+NXA)^{dM}$ values. But $X+1 > (NA)^\delta$, which implies that

$$(X+1)^N = (X+1)^{N-dM}(X+1)^{dM} > (NA)^{dM}(X+1)^{dM} \geq (NAX+1)^{dM}.$$

The pigeonhole principle says that there are distinct N-tuples of integers x' and x'' satisfying $L(x') = L(x'')$. Hence

$$L(x'-x'') = 0 \quad \text{and} \quad |x'-x''| \leq X \leq (NA)^\delta$$

as required. □

We are now ready to construct a polynomial $P(X_1,\dots,X_m)$ that vanishes to high order at $(\alpha,\dots,\alpha)$.

Proposition D.4.3. *Let α be an algebraic integer of degree d over $\mathbb{Q}$, let $\varepsilon > 0$ be a fixed constant, and let m be an integer satisfying*

$$e^{\varepsilon^2 m/4} > 2d. \tag{4.2-i}$$

Let $r_1,\dots,r_m$ be given positive integers. Then there exists a polynomial

$$P(X_1,\dots,X_m) \in \mathbb{Z}[X_1,\dots,X_m]$$

satisfying the following three conditions:

(i) *P has degree at most r_h in the variable X_h.*

(ii) *The index of P with respect to $(\alpha, \alpha, \ldots, \alpha; r_1, \ldots, r_m)$ satisfies*

$$\operatorname{Ind}(P) \geq \frac{m}{2}(1-\varepsilon). \tag{4.2-ii}$$

(iii) *The largest coefficient of P satisfies*

$$|P| \leq B^{r_1+\cdots+r_m}, \qquad \textit{where } B = B(\alpha) \textit{ depends only on } \alpha, \tag{4.2-iii}$$

and we recall that $|P|$ is the maximum of the absolute values of the coefficients of P.

Remark. Condition (4.2-i) essentially ensures that $dM < N$, or more precisely, that $dM/(N-dM) < 1$. An admissible value for the constant given by the proposition is $B(\alpha) = 4H(\alpha)$ (see Remark D.4.4).

PROOF. We write the polynomial P as

$$P(X_1, \ldots, X_m) = \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} p_{j_1,\ldots,j_m} X_1^{j_1} \cdots X_m^{j_m},$$

where the integers $p_{j_1,\ldots,j_m}$ are still to be determined. Clearly, the number of $p_{j_1,\ldots,j_m}$ coefficients is

$$N = (r_1+1)\cdots(r_m+1).$$

For any m-tuple $(i_1, \ldots, i_m)$ we have

$$P_{i_1\ldots i_m} = \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} p_{j_1,\ldots,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} X_1^{j_1-i_1} \cdots X_m^{j_m-i_m}.$$

Evaluating this identity at $(\alpha, \ldots, \alpha)$ and using Lemma D.3.4 to express powers of α as linear combinations of $1, \alpha, \ldots, \alpha^{d-1}$, we find that

$$\begin{aligned}
&P_{i_1\ldots i_m}(\alpha, \ldots, \alpha) \\
&= \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} p_{j_1,\ldots,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} \alpha^{j_1-i_1+\cdots+j_m-i_m} \\
&= \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} p_{j_1,\ldots,j_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} \sum_{k=1}^{d} a_k^{(j_1+\cdots+j_m-i_1-\cdots-i_m)} \alpha^{d-k} \\
&= \sum_{k=1}^{d} \left\{ \sum_{j_1=0}^{r_1} \cdots \sum_{j_m=0}^{r_m} \binom{j_1}{i_1} \cdots \binom{j_m}{i_m} a_k^{(j_1+\cdots+j_m-i_1-\cdots-i_m)} p_{j_1,\ldots,j_m} \right\} \alpha^{d-k}.
\end{aligned}$$

Hence $P_{i_1\dots i_m}(\alpha,\dots,\alpha)$ will equal 0 if each of the quantities in braces equals 0. In other words, we will have $P_{i_1\dots i_m}(\alpha,\dots,\alpha) = 0$ if we choose the $p_{j_1,\dots,j_m}$ to satisfy the d linear equations

$$\sum_{j_1=0}^{r_1}\cdots\sum_{j_m=0}^{r_m}\binom{j_1}{i_1}\cdots\binom{j_m}{i_m}a_k^{(j_1+\cdots+j_m-i_1-\cdots-i_m)}p_{j_1,\dots,j_m} = 0, \qquad 1 \le k \le d.$$

In order to satisfy condition (ii), we need $P_{i_1\dots i_m}(\alpha,\dots,\alpha) = 0$ for all m-tuples $(i_1,\dots,i_m)$ satisfying

$$\frac{i_1}{r_1}+\cdots+\frac{i_m}{r_m} \le \frac{m}{2}(1-\varepsilon) = \frac{m}{2} - \frac{\varepsilon}{2}m.$$

According to (D.3.6), there are at most $(r_1+1)\cdots(r_m+1)e^{-\varepsilon^2 m/4}$ such m-tuples. Hence we can find a P that satisfies (ii) by choosing the $p_{j_1,\dots,j_m}$ to satisfy a system of M linear equations with integer coefficients, where

$$M \le d\cdot(r_1+1)\cdots(r_m+1)e^{-\varepsilon^2 m/4} = dNe^{-\varepsilon^2 m/4} \le \frac{1}{2}N.$$

Note that the last inequality follows from the choice of m in equation (4.2-i).

We now have M linear equations for the N variables $p_{j_1,\dots,j_m}$. In order to apply Siegel's lemma, we need to estimate the size of the coefficients of these equations. We recall from (D.3.4) that the quantities $a_k^{(\ell)}$ satisfy $|a_k^{(\ell)}| \le \bigl(|Q|+1\bigr)^\ell$, where Q is the minimal polynomial for α over $\mathbb{Q}$. So we can estimate the coefficients of our linear equations by

$$\begin{aligned}\left|\binom{j_1}{i_1}\cdots\binom{j_m}{i_m}a_k^{j_1+\cdots+j_m-i_1-\cdots-i_m}\right| &\le 2^{j_1+\cdots+j_m}\bigl(|Q|+1\bigr)^{j_1+\cdots+j_m}\\ &\le \bigl(2|Q|+2\bigr)^{r_1+\cdots+r_m}.\end{aligned}$$

Now applying Siegel's lemma (D.4.1), we find that there is a polynomial P satisfying (i) and (ii) whose coefficients $p_{j_1,\dots,j_m}$ are bounded by

$$\begin{aligned}|P| &\le \Bigl(N\bigl(2|Q|+2\bigr)^{r_1+\cdots+r_m}\Bigr)^{M/(N-M)}\\ &\le N\bigl(2|Q|+2\bigr)^{r_1+\cdots+r_m} \qquad \text{since } M \le \tfrac{1}{2}N\\ &\le 2^{r_1+\cdots+r_m}\bigl(2|Q|+2\bigr)^{r_1+\cdots+r_m}\\ &\le B(\alpha)^{r_1+\cdots+r_m},\end{aligned}$$

where we could take $B(\alpha) = 4|Q|+4$, for example. Hence P also satisfies (iii), which completes the proof of Proposition D.4.3.

□

Remark D.4.4. We outline an alternative proof of Proposition D.4.3 using Lemma D.4.2 instead of Lemma D.4.1. Letting M denote the number of equations with coefficients in $\mathbb{Q}(\alpha)$ and $N=(r_1+1)\cdots(r_m+1)$ the number of unknowns, we get $dM/(N-dM)\le 1$ and

$$H\left(\ldots,\binom{j_1}{i_1}\cdots\binom{j_m}{i_m}\alpha^{\Sigma j_t-i_t},\ldots\right)\le (2H(\alpha))^{r_1+\cdots+r_m}.$$

Hence Lemma D.4.2 gives us a nonzero polynomial with the required degrees and coefficients bounded by

$$(r_1+1)\cdots(r_m+1)\,(2H(\alpha))^{r_1+\cdots+r_m}\le (4H(\alpha))^{r_1+\cdots+r_m},$$

which yields an admissible value $B(\alpha)=4H(\alpha)$.

D.5. The Index Is Large

In the last section we constructed a polynomial with "small" coefficients that vanishes to high order at $(\alpha,\ldots,\alpha)$. In this section we want to show that if $\beta_1,\ldots,\beta_m$ are close to α, then the polynomial will still vanish to high order at $(\beta_1,\ldots,\beta_m)$.

We illustrate the basic idea using polynomials of one variable with $\beta=p/q\in\mathbb{Q}$. Thus suppose that $P(X)\in\mathbb{Z}[X]$ satisfies

$$\deg P(X)=r,\qquad \operatorname{Ind}_{(\alpha;r)}P\ge\frac{1}{2},\qquad |P|\le B(\alpha)^r,$$

and suppose that $p/q\in\mathbb{Q}$ satisfies $|p/q-\alpha|\le q^{-(2+\varepsilon)}$. The Taylor series expansion of P around α is

$$P(X)=\sum_{i=0}^{r}\partial_i P(\alpha)(X-\alpha)^i=\sum_{i\ge r/2}\partial_i P(\alpha)(X-\alpha)^i,$$

since the assumption that $\operatorname{Ind}_{(\alpha;r)}P\ge\frac{1}{2}$ means that $\partial_i P(\alpha)=0$ for all $i<r/2$. Now substituting $X=p/q$ yields

$$\left|P\left(\frac{p}{q}\right)\right|\le\sum_{i\ge r/2}|\partial_i P(\alpha)|\left|\frac{p}{q}-\alpha\right|^i.$$

Using (D.3.1) and $|P|\le B^r$, we can estimate the derivatives by

$$|\partial_i P(\alpha)|\le r\,|\partial_i P|\max\{1,|\alpha|\}^r\le 2^r\cdot B(\alpha)^r\max\{1,|\alpha|\}^r=C_1(\alpha)^r,$$

where the constant $C_1(\alpha)$ depends only on α. Hence

$$\left|P\left(\frac{p}{q}\right)\right| \le r \cdot C_1^r \left|\frac{p}{q} - \alpha\right|^{r/2} \le (2C_1)^r \left(\frac{1}{q^{2+\varepsilon}}\right)^{r/2} = \left(\frac{2C_1}{q^{1+\varepsilon/2}}\right)^r.$$

On the other hand, we know that the denominator of $P(p/q)$ divides q^r, so if P does not vanish at p/q, then trivially $|P(p/q)| \ge 1/q^r$. Hence

$$P\left(\frac{p}{q}\right) \ne 0 \implies \frac{1}{q^r} \le P\left(\frac{p}{q}\right) \le \left(\frac{2C_1}{q^{1+\varepsilon/2}}\right)^r,$$

which implies that $q \le (2C_1)^{2/\varepsilon}$. Turning this around, we can say that if q is large, then $P(p/q) = 0$, so the index of P at $(p/q; r)$ is strictly positive. A similar argument, using $\partial_j P$ in place of P, would show that if q is sufficiently large, then $\partial_j P(p/q) = 0$, or equivalently that the index of P at $(p/q; r)$ is fairly large.

The main result of this section generalizes this simple argument in a number of ways, but the entire computation rests on two basic facts. First, Taylor's formula and the triangle inequality imply that if P vanishes to a high order at α and if β is very close to α, then $P(\beta)$ (and any derivative of not too large an order) must be small. Second, Liouville's inequality (as recorded in Lemma D.3.3) implies that $P(\beta)$, an algebraic number of bounded height, cannot be too small unless it is zero. (Keep in mind that this second fact is essentially nothing more than the observation that there are no integers strictly between 0 and 1.)

Proposition D.5.1. *Let $0 < \delta < 1$ be a given constant, and let ε satisfy*

$$0 < \varepsilon < \frac{\delta}{22}. \tag{5.1-i}$$

Let α be an algebraic integer of degree d over $\mathbb{Q}$, let m be an integer satisfying $e^{\varepsilon^2 m/4} > 2d$, let $r_1, \dots, r_m$ be given positive integers, and use Proposition D.4.3 to choose a polynomial $P(X_1, \dots, X_m) \in \mathbb{Z}[X_1, \dots, X_m]$ satisfying properties (4.2-i), (4.2-ii), and (4.2-iii).

Let $S \subset M_K$ be a finite set of absolute values on K with each absolute value extended in some way to $\bar{K}$, and let

$$\xi : S \to [0,1] \quad \textit{be a function satisfying} \quad \sum_{v \in S} \xi_v = 1.$$

Suppose that $\beta_1, \dots, \beta_m \in K$ have the property that

$$\|\beta_h - \alpha\|_v \le \frac{1}{H_K(\beta_h)^{(2+\delta)\xi_v}} \quad \textit{for all } v \in S \textit{ and all } 1 \le h \le m. \tag{5.1-ii}$$

Suppose further that

$$\max_{1\le h\le m}\{H_K(\beta_h)^{r_h}\} \le \min_{1\le h\le m}\{H_K(\beta_h)^{r_h}\}^{1+\varepsilon} \tag{5.1-iii}$$

and that there is a constant $C = C(\alpha,\delta)$ *such that*

$$C \le H(\beta_h) \quad \text{for all } 1 \le h \le m. \tag{5.1-iv}$$

Then the index of P *with respect to* $(\beta_1,\dots,\beta_m; r_1,\dots,r_m)$ *satisfies*

$$\operatorname{Ind} P \ge \varepsilon m.$$

Remark D.5.1.1. When we eventually get to the proof of Roth's theorem in Section 7, the proof will proceed as follows. We will first assume that there are infinitely many β's that closely approximate α, that is, satisfying the inequality (5.1-ii). Next we will choose a large m and good approximations $\beta_1,\dots,\beta_m$ such that each β_h has height much larger than its predecessor. It is only after choosing the β_h's that we will choose the r_i's so as to satisfy the estimate (5.1-iii). Finally, for this choice of m and $r_1,\dots,r_m$, we will take the auxiliary polynomial described in Proposition D.4.3 and use it to prove Roth's theorem. The crucial observation here is that the r_i's are chosen in terms of the hypothetical β_h's, so it is important at every step of the proof to keep track of any dependence on the r_i's.

As a first step toward the proof of Proposition D.5.1, we state a lemma controlling the height of derivatives of a polynomial at algebraic values. Note that this result and its proof are very similar to Theorem B.2.5(a).

Lemma D.5.2. *Let* $P \in \mathbb{Z}[X_1,\dots,X_m]$ *with* $\deg_{X_h}(P) \le r_h$, *and let* $\beta = (\beta_1,\dots,\beta_m)$ *be an* m*-tuple of algebraic numbers in a number field* K. *Then for all* m*-tuples of nonnegative integers* $j = (j_1,\dots,j_m)$ *we have*

$$H_K\big(\partial_j P(\beta)\big) \le 4^{(r_1+\cdots+r_m)[K:\mathbb{Q}]} H_K(P) \prod_{h=1}^{m} H_K(\beta_h)^{r_h}.$$

PROOF. Let $(j_1,\dots,j_m)$ be any m-tuple. To simplify notation, we will let

$$T(X_1,\dots,X_m) = \partial_{j_1\dots j_m} P(X_1,\dots,X_m).$$

Then Lemma D.3.1 tells us that $\partial_{i_1\dots i_m} T$ has integer coefficients that are bounded by

$$|\partial_{i_1\dots i_m} T| = |\partial_{i_1+j_1,\dots,i_m+j_m} P| \le 2^{r_1+\cdots+r_m}|P|.$$

We use the triangle inequality to find an upper bound for the height of $T(\beta_1,\dots,\beta_m)$. Thus for any archimedean absolute value $v \in M_K^\infty$ we have

$$\begin{aligned}
&\bigl|T(\beta_1,\dots,\beta_m)\bigr|_v \\
&\quad\le \underbrace{(r_1+1)\cdots(r_m+1)}_{\text{number of terms}}\cdot \underbrace{|T|}_{\text{max. coef.}} \cdot \max\bigl\{|\beta_1|_v,1\bigr\}^{r_1}\cdots\max\bigl\{|\beta_m|_v,1\bigr\}^{r_m} \\
&\quad\le (4)^{r_1+\cdots+r_m}|P|\cdot\max\bigl\{|\beta_1|_v,1\bigr\}^{r_1}\cdots\max\bigl\{|\beta_m|_v,1\bigr\}^{r_m}.
\end{aligned}$$

Similarly, if $v \in M_K^0$ is nonarchimedean, we can use the nonarchimedean triangle inequality and the fact that T has integer coefficients to obtain the stronger bound

$$\bigl|T(\beta_1,\dots,\beta_m)\bigr|_v \le \max\bigl\{|\beta_1|_v,1\bigr\}^{r_1}\cdots\max\bigl\{|\beta_m|_v,1\bigr\}^{r_m}.$$

As in the definition of the height, we raise these inequalities to the n_v power (where $n_v = [K_v : \mathbb{Q}_v]$) and multiply them together for all $v \in M_K$. This yields the announced estimate

$$H_K\bigl(T(\beta_1,\dots,\beta_m)\bigr) \le (4)^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}H_K(P)H_K(\beta_1)^{r_1}\cdots H_K(\beta_m)^{r_m}.$$

(Notice that since P has integer coefficients, $H_K(P) = |P|^{[K:\mathbb{Q}]}$.) $\square$

We now give an explicit version of the principle that if a polynomial vanishes to a high order at $(\alpha,\dots,\alpha)$, and if $\beta_1,\dots,\beta_m$ are all good approximations to α, then the polynomial will be very small at $(\beta_1,\dots,\beta_m)$.

Lemma D.5.3. *Let $r_1,\dots,r_m$ be given positive integers, and let P be a polynomial in $\mathbb{Z}[X_1,\dots,X_m]$ such that $\deg_{X_h}(P) \le r_h$. Let $\theta = \operatorname{Ind} P$ denote the index of P at $(\alpha,\dots,\alpha)$ with respect to $r_1,\dots,r_m$. Let $0<\delta<1$ be a constant and choose $0<\theta_0<\theta$.*

Let $S \subset M_K$ be a finite set of absolute values on K with each absolute value extended in some way to $\bar{K}$, and let

$$\xi : S \to [0,1] \quad \text{be a function satisfying} \quad \sum_{v\in S}\xi_v = 1.$$

Suppose that $\beta_1,\dots,\beta_m \in K$ have the property that

$$\|\beta_h-\alpha\|_v \le \frac{1}{H_K(\beta_h)^{(2+\delta)\xi_v}} \quad \text{for all } v\in S \text{ and all } 1\le h\le m.$$

To ease notation, set $D := \min\{H_K(\beta_h)^{r_h}\}$, and let $j=(j_1,\dots,j_m)$ be any m-tuple satisfying $\displaystyle\sum_{h=1}^{m}\frac{j_h}{r_h}\le\theta_0$. Then

$$\prod_{v\in S}\bigl\|\partial_j P(\beta_1,\dots,\beta_m)\bigr\|_v \le (4H(\alpha))^{[K:\mathbb{Q}](r_1+\cdots+r_m)}H_K(P)D^{-(2+\delta)(\theta-\theta_0)}.$$

PROOF. Let $j = (j_1, \dots, j_m)$ be as in the statement, and let $T = \partial_j P$. We will use the Taylor expansion of T around $(\alpha, \dots, \alpha)$ and will need bounds for the size of the Taylor coefficients. Thus if $v \in M_K$ is an absolute value on K extended in some way to $K(\alpha)$, we can estimate $\left|\partial_{i_1\dots i_m} T(\alpha, \dots, \alpha)\right|_v$ by noting that it is a sum of at most

$$(r_1 + 1) \cdots (r_m + 1)$$

terms, each of which has magnitude at most

$$|T| \max\bigl\{|\alpha|_v, 1\bigr\}^{r_1 + \cdots + r_m} \le |P| \bigl(2 \max\bigl\{|\alpha|_v, 1\bigr\}\bigr)^{r_1 + \cdots + r_m}.$$

This implies that

$$\left|\partial_{i_1\dots i_m} T(\alpha, \dots, \alpha)\right|_v \le \bigl(4 \max\bigl\{|\alpha|_v, 1\bigr\}\bigr)^{r_1 + \cdots + r_m} |P|.$$

Next we observe that T vanishes to fairly high order at $(\alpha, \dots, \alpha)$. More precisely, we use Lemma D.3.2(a) to compute

$$\operatorname{Ind} T = \operatorname{Ind} \partial_{j_1 \dots j_m} P \ge \operatorname{Ind} P - \sum_{h=1}^{m} \frac{j_h}{r_h} \ge \theta - \theta_0.$$

So if we write the Taylor expansion of T about $(\alpha, \dots, \alpha)$, then many of the initial terms will be zero. Thus

$$T(X_1, \dots, X_m) = \sum_{\substack{i_1 = 0 \\ \frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m} \ge \theta - \theta_0}}^{r_1} \cdots \sum_{i_m = 0}^{r_m} \partial_{i_1 \dots i_m} T(\alpha, \dots, \alpha) (X_1 - \alpha)^{i_1} \cdots (X_m - \alpha)^{i_m}.$$

Now put $X_h = \beta_h$ and use the fact that β_h is close to α. This yields, for each absolute value $v \in S$,

$$\begin{aligned}
&\left|T(\beta_1, \dots, \beta_m)\right|_v \\
&\quad \le \sum_{\substack{i_1 = 0 \\ \frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m} \ge \theta - \theta_0}}^{r_1} \cdots \sum_{i_m = 0}^{r_m} \left|\partial_{i_1 \dots i_m} T(\alpha, \dots, \alpha)\right|_v |\beta_1 - \alpha|_v^{i_1} \cdots |\beta_m - \alpha|_v^{i_m} \\
&\quad \le (r_1 + 1) \cdots (r_m + 1) \max_{i_1, \dots, i_m} \left|\partial_{i_1 \dots i_m} T(\alpha, \dots, \alpha)\right|_v \\
&\qquad\qquad \times \max_{\frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m} \ge \theta - \theta_0} |\beta_1 - \alpha|_v^{i_1} \cdots |\beta_m - \alpha|_v^{i_m} \\
&\quad \le (4)^{r_1 + \cdots + r_m} |P|_v \max_{\frac{i_1}{r_1} + \cdots + \frac{i_m}{r_m} \ge \theta - \theta_0} \frac{1}{\bigl(H_K(\beta_1)^{i_1} \cdots H_K(\beta_m)^{i_m}\bigr)^{(2+\delta)\xi_v}}.
\end{aligned}$$

We can estimate this last quantity as follows:

$$H_K(\beta_1)^{i_1}\cdots H_K(\beta_m)^{i_m} = (H_K(\beta_1)^{r_1})^{\frac{i_1}{r_1}}\cdots(H_K(\beta_m)^{r_m})^{\frac{i_m}{r_m}} \geq D^{\theta-\theta_0}$$

It follows from above that

$$\left|T(\beta_1,\ldots,\beta_m)\right|_v \leq \frac{(4)^{r_1+\cdots+r_m}|P|_v}{D^{(\theta-\theta_0)(2+\delta)\xi_v}}.$$

Now raising to the n_v power, multiplying over all $v\in S$, and using the fact that $\sum_{v\in S} n_v\xi_v \geq \sum_{v\in S}\xi_v = 1$, we arrive at the desired estimate

$$\prod_{v\in S}\left\|T(\beta_1,\ldots,\beta_m)\right\|_v \leq \frac{4^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}H_K(P)}{D^{(\theta-\theta_0)(2+\delta)}}.$$

□

PROOF (of Proposition D.5.1). Let $j=(j_1,\ldots,j_m)$ be an m-tuple satisfying $\sum_{h=1}^m j_h/r_h \leq \varepsilon m$. We want to show that $\partial_j P(\beta_1,\ldots,\beta_m)=0$. From Lemma D.5.3 we get

$$\begin{aligned}\prod_{v\in S}\left\|\partial_j P(\beta_1,\ldots,\beta_m)\right\|_v &\leq \frac{(4)^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}H_K(P)}{D^{(\theta-\theta_0)(2+\delta)}}\\ &\leq \frac{(4B(\alpha))^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}}{D^{\left(\frac{m}{2}(1-\varepsilon)-\varepsilon m\right)(2+\delta)}}\end{aligned}$$

(where we have used properties (4.3-ii) and (4.3-iii) for the last inequality). On the other hand, from Lemma D.5.2 we obtain

$$\begin{aligned}H_K(\partial_j P(\beta_1,\ldots,\beta_m) &\leq 4^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}H_K(P)\prod_{h=1}^m H_K(\beta_h)^{r_h}\\ &\leq \left(4B(\alpha)\right)^{(r_1+\cdots+r_m)[K:\mathbb{Q}]}D^{m(1+\varepsilon)}\end{aligned}$$

(where we used property (4.3-iii) and hypothesis (5.1-iii) for the last inequality). Now Liouville's inequality (D.3.3) implies that either the derivative $\partial_j P(\beta_1,\ldots,\beta_m)$ is zero, or else

$$\prod_{v\in S}\left\|\partial_j P(\beta_1,\ldots,\beta_m)\right\|_v \geq H_K(\partial_j P(\beta_1,\ldots,\beta_m))^{-1}.$$

So it suffices to show that our hypotheses contradict the latter.

Assuming $\partial_j P(\beta_1,\ldots,\beta_m)\neq 0$, Liouville's inequality thus yields

$$D^{m((1+\delta/2)(1-3\varepsilon)-(1+\varepsilon))} \leq (4B(\alpha))^{2(r_1+\cdots+r_m)[K:\mathbb{Q}]}.$$

Now, since we assumed $\delta<1$ and $\varepsilon<\delta/22$, we get

$$(1+\delta/2)(1-3\varepsilon)-(1+\varepsilon) < \delta/2 - 11\varepsilon/2 < \delta/4,$$

and hence

$$\max_{1\le j\le m}\{H_K(\beta_h)^{r_h}\} \le D^{1+\varepsilon} \le \bigl(4B(\alpha)\bigr)^{8(r_1+\cdots+r_m)[K:\mathbb{Q}](1+\varepsilon)/\delta}.$$

Selecting j such that $r_j = \max_{h=1}^m r_h$, we deduce that

$$H_K(\beta_j) \le \bigl(4B(\alpha)\bigr)^{8[K:\mathbb{Q}](1+\varepsilon)/\delta}.$$

Choosing the constant $C = C(\alpha,\delta)$ of hypothesis (5.1-iv) sufficiently large (for example, $C(\alpha,\delta) = \bigl(4B(\alpha)\bigr)^{(8/\delta)+4/11)}$ suffices), we obtain the desired contradiction, which concludes the proof of Proposition D.5.1. □

D.6. The Index Is Small (Roth's Lemma)

In the last section we showed that the polynomial P, which vanishes to high order at $(\alpha,\ldots,\alpha)$, also vanishes to fairly high order at $(\beta_1,\ldots,\beta_m)$. We now want to show that it is actually not possible for P to vanish to high order at $(\beta_1,\ldots,\beta_m)$, which will give a contradiction to the existence of infinitely many close approximations to α.

For a polynomial of one variable, the idea is very simple. Suppose that the polynomial $P(X)\in\mathbb{Z}[X]$ has size bounded by $|P|\le B^r$, and let I denote the index of P with respect to $(p/q;r)$ for some rational number $p/q\in\mathbb{Q}$. Then

$$\begin{aligned}
\partial_i P\left(\frac{p}{q}\right) = 0 \quad & \text{for all } i/r \le I, \text{ by definition of the index,}\\
\Longrightarrow \left(X-\frac{p}{q}\right)^{rI} & \Bigm| P(X)\\
\Longrightarrow (qX-p)^{rI} \mid P(X) \quad & \text{Gauss's lemma}\\
\Longrightarrow \max\bigl\{|p|,|q|\bigr\}^{rI} \le |P| \le B^r \quad & \text{since } q^{rI} \text{ divides the leading}\\
& \text{coefficient of } P, \text{ and } p^{rI} \text{ divides the constant term of } P\\
\Longrightarrow \operatorname{Ind} P = I \le \frac{\log B}{\log H(p/q)}. &
\end{aligned}$$

Hence $\operatorname{Ind} P$ will be small if both $H(p/q)$ and r are large. Notice that one of the key steps is Gauss's lemma (see Lang [2, V, Corollary 6.2] or Herstein [1, Theorem 3.10.1]), which asserts that if a polynomial with integer coefficients factors in $\mathbb{Q}[X]$, then it factors in $\mathbb{Z}[X]$. We will use a related, but more precise, result (Gelfand's inequality (B.7.3)) in our proof of Roth's lemma.

Unfortunately, using a polynomial of one variable suffices to prove only the elementary estimate of Liouville. And as soon as the polynomial contains more than one variable, the simple divisibility argument given above no longer works. Thue [1] worked with a polynomial of the form $P(X,Y) = f(X) + g(X)Y$. As preparation for the proof of Roth's lemma, we will briefly sketch Thue's idea, which is to eliminate the variable Y and thereby reduce to the one-variable case.

Thus suppose that $P(X,Y) = f(X) + g(X)Y$ has index $I = \operatorname{Ind} P$ with respect to $(\beta_1, \beta_2; r, 1)$. This means that

$$\partial_{i,0}P(\beta_1,\beta_2) = \partial_i f(\beta_1) + \partial_i g(\beta_1)\beta_2 = 0 \qquad \text{for all } i/r \le I.$$

We now consider the Wronskian determinant, which is the polynomial in one variable defined by

$$W(X) = \det \begin{vmatrix} f(X) & g(X) \\ \partial_1 f(X) & \partial_1 g(X) \end{vmatrix} = f(X)\partial_1 g(X) - g(X)\partial_1 f(X).$$

Differentiating the Wronskian determinant k times, one easily checks that

$$\partial_k W = \sum_{i+j=k} \big(\partial_i f \cdot \partial_{j+1} g - \partial_{j+1} f \cdot \partial_i g\big).$$

On the other hand, we know from above that if $i \le rI$ and $j+1 \le rI$, then

$$0 = \partial_{i,0}P(\beta_1,\beta_2) = \partial_i f(\beta_1) + \partial_i g(\beta_1)\beta_2$$

and

$$0 = \partial_{j+1,0}P(\beta_1,\beta_2) = \partial_{j+1} f(\beta_1) + \partial_{j+1} g(\beta_1)\beta_2.$$

Eliminating β_2 from these two equations gives

$$\partial_i f(\beta_1)\partial_{j+1} g(\beta_1) - \partial_{j+1} f(\beta_1)\partial_i g(\beta_1) = 0 \qquad \text{for all } i \le rI,\ j \le rI - 1.$$

It follows that $\partial_k W(\beta_1) = 0$ for all $k \le rI - 1$, which means that the index of W with respect to $(\beta_1; r)$ satisfies

$$\operatorname{Ind} W \ge \operatorname{Ind} P - \frac{1}{r}.$$

Now one can estimate the size of $|W|$ and use the one-variable argument to get an upper bound for $\operatorname{Ind} W$, thereby obtaining an upper bound for $\operatorname{Ind} P$; this leads to an approximation exponent with value $1 + d/2$.

The proof of the general case of Roth's lemma proceeds similarly by induction on the number of variables. Thus starting with a polynomial $P(X_1, \ldots, X_m)$ in m variables, we take a determinant of derivatives to form a new polynomial $W(X_1, \ldots, X_m)$ that factors as a product of

the form $V(X_1,\ldots,X_{m-1})U(X_m)$. Lemma D.3.2(c) tells us that $\operatorname{Ind} W = \operatorname{Ind} V + \operatorname{Ind} U$, and then applying the inductive hypotheses to U and V gives the desired bound. Needless to say, the estimates needed to make this induction work are very delicate, and this brief summary has omitted a number of crucial details, which we now begin to fill in.

Let $f_1(X),\ldots,f_n(X) \in K(X)$ be rational functions of a single variable. The *classical Wronskian determinant* of $f_1(X),\ldots,f_n(X)$ is the function

$$W(f_1,\ldots,f_n) = \det\begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ \dfrac{df_1}{dX} & \dfrac{df_2}{dX} & \cdots & \dfrac{df_n}{dX} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{d^{n-1}f_1}{dX^{n-1}} & \dfrac{d^{n-1}f_2}{dX^{n-1}} & \cdots & \dfrac{d^{n-1}f_n}{dX^{n-1}} \end{pmatrix}.$$

It is a standard theorem that the functions $f_1,\ldots,f_n$ are linearly independent over K if and only if $W(f_1,\ldots,f_n) \neq 0$. We will need a version of the Wronskian determinant that applies to polynomials of more than one variable.

Definition. The *order of a differential operator*

$$\Delta = \frac{\partial^{i_1+\cdots+i_m}}{\partial X_1^{i_1}\cdots X_m^{i_m}}$$

is the quantity $\operatorname{order}(\Delta) = i_1 + \cdots + i_m$.

Of course, for arithmetic applications it is often better to use operators of the form

$$\frac{1}{i_1!\cdots i_m!}\frac{\partial^{i_1+\cdots+i_m}}{\partial X_1^{i_1}\cdots X_m^{i_m}}.$$

Now let $\phi_1,\ldots,\phi_k \in K(X)$ be rational functions over a field of characteristic 0. A *generalized Wronskian determinant* of $\phi_1,\ldots,\phi_k$ is any determinant of the form

$$\det\bigl((\Delta_i\phi_j)_{1\le i,j\le k}\bigr),$$

where the differential operators Δ_i satisfy $\operatorname{order}(\Delta_i) \le i-1$.

For example, if $m=1$, then up to constants the only nontrivial generalized Wronskian has $\Delta_i = d^{i-1}/dX^{i-1}$, so will be equal to a constant multiple of the classical Wronskian determinant. The following lemma generalizes the older linear independence result.

Lemma D.6.1. *Let $\phi_1,\ldots,\phi_k \in K(X_1,\ldots,X_m)$ be rational functions over a field of characteristic 0. Then $\phi_1,\ldots,\phi_k$ are linearly independent over K if and only if there exists a nonzero generalized Wronskian of $\phi_1,\ldots,\phi_k$.*

PROOF. We will prove that if $\phi_1, \ldots, \phi_k$ are linearly independent, then there exists a nonzero generalized Wronskian for them. This is the direction that we will need for the proof of Roth's lemma. We leave the proof of the converse as an exercise for the reader.

The proof is by induction on k, the number of functions. We suppose that $\phi_1, \ldots, \phi_k$ are linearly independent. For $k = 1$, the only generalized Wronskian determinant is a constant multiple of $\Delta_1\phi_1 = \phi_1$, since the operator Δ_1 has order zero. So for $k = 1$, Lemma D.6.1 just says that ϕ_1 is linearly independent over K if and only if $\phi_1 \neq 0$.

Assume now that we know that the lemma is true for any set of $k-1$ functions, and suppose that $\phi_1, \ldots, \phi_k$ are K-linearly independent. We need to find a nonzero generalized Wronskian of $\phi_1, \ldots, \phi_k$. We observe that if $\lambda \in K(X_1, \ldots, X_m)$ is a nonzero function, then any generalized Wronskian $\det\big(\Delta_i(\lambda\phi_j)\big)$ of $\lambda\phi_1, \ldots, \lambda\phi_k$ is a $K(X_1, \ldots, X_m)$-linear combination of generalized Wronskians of $\phi_1, \ldots, \phi_k$. This is easily verified using the product rule and the multilinearity of the determinant. So if we can show that some generalized Wronskian of $\lambda\phi_1, \ldots, \lambda\phi_k$ is nonzero, then it will follow that some generalized Wronskian of $\phi_1, \ldots, \phi_k$ is nonzero. Taking $\lambda = 1/\phi_1$, we have reduced to the case that $\phi_1 = 1$.

Let

$$V = K\phi_1 + K\phi_2 + \cdots + K\phi_k \subset K(X_1, \ldots, X_m)$$

be the K-linear span of $\phi_1, \ldots, \phi_k$. By assumption, we know that $\dim V = k$. In particular, $\phi_2 \notin K$ (i.e., ϕ_2 is not a constant function, since $\phi_1 = 1$), so after relabeling the variables, we may assume that the variable X_1 appears in ϕ_2. In other words, we may assume that

$$\frac{\partial\phi_2}{\partial X_1} \neq 0.$$

Define a K-vector subspace of V by

$$W = \left\{\phi \in V \;\middle|\; \frac{\partial\phi}{\partial X_1} = 0\right\}, \quad \text{and let } t = \dim W.$$

We observe that $\phi_1 \in W$ and $\phi_2 \notin W$, so $1 \le t \le k-1$. Now choose a basis $\psi_1, \ldots, \psi_t$ for W, and extend it to a basis $\psi_1, \ldots, \psi_k$ for V. By the induction hypothesis, there are differential operators $\Delta_1^*, \ldots, \Delta_t^*$ satisfying

$$\det(\Delta_i^*\psi_j)_{1\le i,j\le t} \neq 0 \qquad \text{and} \qquad \operatorname{order}(\Delta_i^*) \le i-1.$$

Next we claim that the functions $\partial\psi_{t+1}/\partial X_1, \ldots \partial\psi_k/\partial X_1$ are K-linearly independent. This is true because $\{\psi_{t+1}, \ldots, \psi_k\}$ is a K-basis for the quotient space V/W, so

$$\sum_{i=t+1}^{k} c_i\frac{\partial\psi_i}{\partial X_1} = 0 \Longrightarrow \sum_{i=t+1}^{k} c_i\psi_i \in W \Longrightarrow c_{t+1} = \cdots = c_k = 0.$$

So we can apply the induction hypothesis again to find differential operators $\Delta_{t+1}^*, \ldots, \Delta_k^*$ satisfying

$$\det\left(\Delta_i^* \frac{\partial \psi_j}{\partial X1}\right)_{t+1 \le i,j \le k} \ne 0 \qquad \text{and} \qquad \operatorname{order}(\Delta_i^*) \le i - t - 1.$$

We want to fit these two determinants together, so we define differential operators

$$\Delta_i = \begin{cases} \Delta_i^* & \text{if } 1 \le i \le t, \\ \Delta_i^* \frac{\partial}{\partial X_1} & \text{if } t+1 \le i \le k. \end{cases}$$

Note that $\operatorname{order}(\Delta_i) \le i - 1$ for all $1 \le i \le k$. Further, and this is crucial, we have

$$\Delta_i \psi_j = \Delta_i^* \frac{\partial \psi_j}{\partial X_1} = 0 \quad \text{for all } 1 \le j \le t \text{ and } t+1 \le i \le k.$$

This is true because $\psi_j \in W$ for $1 \le j \le t$. Hence we find that

$$\begin{aligned} \det(\Delta_i \psi_j)_{1 \le i,j \le k} &= \det \begin{pmatrix} \Delta_i^* \psi_j & \Delta_i^* \psi_j \\ 0 & \Delta_i^* \dfrac{\partial \psi_j}{\partial X_1} \end{pmatrix} \begin{matrix} \uparrow t \downarrow \\ \uparrow k-t \downarrow \end{matrix} \\ &\qquad\;\; \overset{\longleftarrow t \longrightarrow \quad \longleftarrow k-t \longrightarrow}{} \\ &= \det(\Delta_i^* \psi_j)_{1 \le i,j \le t} \cdot \det\left(\Delta_i^* \frac{\partial \psi_j}{\partial X_1}\right)_{t+1 \le i,j \le k} \\ &\ne 0 \quad \text{from above.} \end{aligned}$$

This shows that $\psi_1, \ldots, \psi_k$ have a nonzero generalized Wronskian. But the functions $\phi_1, \ldots, \phi_k$ and $\psi_1, \ldots, \psi_k$ span the same K-vector space, so $\psi_j = \sum_\ell a_{j\ell} \phi_\ell$ for some invertible matrix $(a_{j\ell})$ with coefficients in K. It follows that

$$0 \ne \det(\Delta_i \psi_j) = \det\left(\sum_\ell a_{j\ell} \Delta_i \phi_\ell\right) = \det(a_{j\ell}) \det(\Delta_i \phi_\ell),$$

and so $\det(\Delta_i \phi_\ell) \ne 0$. This concludes the proof of Lemma D.6.1. □

We are now ready to begin the proof of Roth's lemma, which says that the polynomial $P(X_1, \ldots, X_m)$ constructed earlier does not vanish to high order at $(\beta_1, \ldots, \beta_m)$. Although the proof is fairly lengthy, the essential idea is to use Wronskians to eliminate a variable and then perform an induction on the number of variables.

Proposition D.6.2. (Roth's lemma) *Let m be a positive integer and let $P \in \bar{\mathbb{Q}}[X_1, \dots, X_m]$ be a polynomial with algebraic coefficients and $\deg_{X_h}(P) \le r_h$. Let $\beta = (\beta_1, \dots, \beta_m)$ be an m-tuple of algebraic numbers. Fix a real number $0 < \eta$ such that*

$$\frac{r_{h+1}}{r_h} \le \eta^{2^{m-1}} \quad \text{for all } 1 \le h \le m-1, \tag{6.2-i}$$

and

$$\eta^{2^{m-1}} \min_{1 \le h \le m} \{r_h \log H(\beta_h)\} \ge \log H(P) + 2mr_1. \tag{6.2-ii}$$

Then the index of P with respect to $(\beta_1, \dots, \beta_m; r_1, \dots, r_m)$ satisfies

$$\operatorname{Ind} P \le 2m\eta.$$

Remark. In practice, η will be very small. (Indeed, if $\eta \ge \frac{1}{2}$, then the conclusion of the proposition is trivial, since we always have $\operatorname{Ind} P \le m$.) It follows from (6.2-i) that the degrees $r_1, r_2, \dots, r_m$ are very rapidly decreasing, and then (6.2-ii) implies that the heights $H(\beta_1), \dots, H(\beta_m)$ are very rapidly increasing. It is useful to keep these two properties in mind.

PROOF (of Roth's lemma). The proof is by induction on m, the number of variables. To ease notation, for the remainder of this section we will let K be a number field containing all the β_i's and the coefficients of P. We also let

$$d = [K : \mathbb{Q}].$$

We also recall that the height of a polynomial is defined to be the height of its coefficients (B.7). So for example, if a polynomial F has coefficients in $\mathbb{Z}$, then $H_K(F) = |F|^d$.

We begin with the case $m = 1$. To ease notation, we write $\beta = \beta_1$ and $r = r_1$. Let ℓ be the exact order of vanishing of $P(X)$ at $X = \beta$, so $P(X) = (X - \beta)^\ell Q(X)$ with $Q(\beta) \ne 0$. Note that the index of P at $(\beta; r)$ is then $\operatorname{Ind} P = \ell/r$. Using Gelfand's inequality (B.7.3), we find that

$$H(\beta)^{r \operatorname{Ind} P} = H(\beta)^\ell = H(X - \beta)^\ell \le H(X - \beta)^\ell H(Q) \le H(P) e^r,$$

which implies

$$\operatorname{Ind} P \le \frac{\log H(P) + r}{r \log H(\beta)} \le \eta \qquad \text{using hypothesis (6.2-ii).}$$

This completes the proof of Roth's lemma for polynomials of one variable.

Remark. This is slightly better than the stated result. We get an upper bound for the index equal to η, instead of 2η, and we needed only $\eta r \log H(\beta) \geq \log H(P) + r$ instead of $\geq \log H(P) + 2r$. Later we will make use of this observation to slightly sharpen our estimates when using induction with $m = 1$.

We now assume that Roth's lemma is true for polynomials with strictly fewer than m variables, and we prove it for a polynomial $P(X_1, \ldots, X_m)$ of m variables. We begin by writing P in the form

$$P(X_1, \ldots, X_m) = \sum_{j=1}^{k} \phi_j(X_1, \ldots, X_{m-1}) \psi_j(X_m), \tag{$*$}$$

where the ϕ_j's and ψ_j's are polynomials with coefficients in $\bar{\mathbb{Q}}$. There are many such ways to decompose P, and among them, we choose one for which k is smallest. That is, we choose a decomposition $(*)$ with the smallest number of summands. Since one possible decomposition (probably not minimal) is to take $\psi_1 = 1$, $\psi_2 = X_m$, $\psi_3 = X_m^2$, $\ldots \psi_k = X_m^{r_m}$, we see that

$$k \leq r_m + 1.$$

Claim D.6.2.1. *The functions $\phi_1, \ldots, \phi_k$ appearing in the minimal decomposition $(*)$ of P (described above) are $\bar{\mathbb{Q}}$-linearly independent. Similarly, the functions $\psi_1, \ldots, \psi_k$ are $\bar{\mathbb{Q}}$-linearly independent.*

PROOF (of Claim). Suppose that the ϕ_j's are linearly dependent, so there is a nontrivial linear relation $\sum c_j \phi_j = 0$. Relabeling if necessary, we may assume that $c_k \neq 0$. Then

$$\phi_k = -\sum_{j=1}^{k-1} \frac{c_j}{c_k} \phi_j,$$

and so

$$P = \sum_{j=1}^{k} \phi_j \psi_j = \sum_{j=1}^{k-1} \phi_j \psi_j - \sum_{j=1}^{k-1} \frac{c_j}{c_k} \phi_j \psi_j = \sum_{j=1}^{k-1} \phi_j \left(\psi_j - \frac{c_j}{c_k} \psi_k \right),$$

contradicting the minimality of k. This proves that $\phi_1, \ldots, \phi_k$ are linearly independent, and the proof for $\psi_1, \ldots, \psi_k$ is the same. This concludes the proof of the claim. □

Define a polynomial $U(X_m)$ by

$$U(X_m) \stackrel{\text{def}}{=} \det\left(\frac{1}{(i-1)!} \frac{\partial^{i-1}}{\partial X_m^{i-1}} \psi_j(X_m) \right)_{1 \leq i,j \leq k}.$$

This is the (classical) Wronskian determinant of $\psi_1, \dots, \psi_k$, so we know from Lemma D.6.1 and the claim that $U(X_m) \neq 0$. Similarly, Lemma D.6.1 and the independence of $\phi_1, \dots, \phi_k$ imply that we can find differential operators

$$\Delta_i' = \frac{1}{i_1! \cdots i_{m-1}!} \frac{\partial^{i_1+\cdots+i_m}}{\partial X_1^{i_1} \cdots X_{m-1}^{i_{m-1}}}$$

with

$$\operatorname{order}(\Delta_i') = i_1 + \cdots + i_{m-1} \le i - 1 \le k - 1 \le r_m$$

such that the generalized Wronskian determinant satisfies

$$V(X_1, \dots, X_{m-1}) \stackrel{\text{def}}{=} \det(\Delta_i' \phi_j)_{1 \le i,j \le k} \neq 0.$$

We now define a polynomial $W(X_1, \dots, X_m)$ and use the fact that the differential operators Δ_i' involve only $X_1, \dots, X_{m-1}$ to compute

$$\begin{aligned}
W(X_1, \dots, X_m) &\stackrel{\text{def}}{=} \det\left(\Delta_i' \left(\frac{1}{(j-1)!} \frac{\partial^{j-1}}{\partial X_m^{j-1}}\right) P(X_1, \dots, X_m)\right)_{1 \le i,j \le k} \\
&= \det\left(\Delta_i' \left(\frac{1}{(j-1)!} \frac{\partial^{j-1}}{\partial X_m^{j-1}}\right) \sum_{r=1}^{k} \phi_r(X_1, \dots, X_{m-1}\psi_r(X_m)\right)_{1 \le i,j \le k} \\
&= \det\left(\sum_{r=1}^{k} \Delta_i' \phi_r \cdot \frac{1}{(j-1)!} \frac{\partial^{j-1}\psi_r}{\partial X_m^{j-1}}\right)_{1 \le i,j \le k} \\
&= \det(\Delta_i' \phi_r)_{1 \le i,r \le k} \cdot \det\left(\frac{1}{(j-1)!} \frac{\partial^{j-1}\psi_r}{\partial X_m^{j-1}}\right)_{1 \le j,r \le k} \\
&\qquad\qquad \text{(matrix multiplication!)} \\
&= V(X_1, \dots, X_{m-1}) U(X_m).
\end{aligned}$$

Thus the use of the Wronskian determinants allows us to create a polynomial W that is closely related to P and that factors into two polynomials each involving fewer variables than P. We also observe that $W \in K[X_1, \dots, X_m]$. The remainder of the proof of Roth's lemma consists of two basic steps.

[i] Use induction, more precisely Roth's lemma in 1 and $m-1$ variables, to get an upper bound for the indices of U and V, and use these values to get an upper bound for the index of W.

[ii] Relate the index of W back to the index of P. More precisely, write a lower bound for the index of W in terms of the index of P.

We have already remarked that we may assume $\eta \le \frac{1}{2}$. We also note that since U and V use disjoint sets of variables, it is clear from the definition of height of a polynomial that

$$h(U) + h(V) = h(W).$$

In order to apply Roth's lemma to U and V (with 1 or $m-1$ variables, respectively), we need estimates for their degrees and heights.

Claim D.6.2.2. *The following estimates are valid:*
(a) $\deg_{X_m}(U) \le kr_m$ *and* $\deg_{X_j}(V) \le kr_j$ *for all* $1 \le j \le m-1$.
(b) $h(U) + h(V) = h(W) \le k\big(h(P) + 2r_1\big)$.

PROOF (of Claim). (a) Each determinant is of size k and the entries of V (respectively U) have degree at most r_j with respect to X_j (respectively at most r_m with respect to X_m).
(b) The determinant is the sum of $k!$ terms, each of which is a product of k polynomials of degree at most r_j with respect to X_j and satisfying

$$H(\Delta_i'\partial_j P(X_1,\dots,X_m)) \le 2^{r_1+\cdots+r_m} H(P).$$

Thus using (B.7.2), we obtain a bound

$$h(W) \le k\,(h(P) + (r_1 + \cdots + r_m)\log 2) + \log(k!).$$

Now

$$r_1 + \cdots + r_m \le r_1(1 + \eta' + \cdots + \eta'^{m-1}) \quad \text{with} \quad \eta' = \eta^{2^{m-1}}.$$

Since $\eta \le \frac{1}{2}$ and $m \ge 2$, we have $\eta' \le \frac{1}{4}$ and $r_1 + \cdots + r_m \le \frac{4}{3} r_1$. On the other hand,

$$\frac{\log(k!)}{k} \le \log(k) \le k - 1 \le r_m \le \frac{1}{2} r_1,$$

and hence

$$h(W) \le k\left(h(P) + \left(\frac{4}{3}\log 2 + \frac{1}{2}\right) r_1\right) \le k\,(h(P) + 2r_1)\,.$$

We note for future reference that the constant 2 (in front of r_1 in the last inequality of the claim) could be replaced by the smaller constant $c_1 = \frac{4}{3}\log 2 + \frac{1}{2} \approx 1.424$. □

We now use induction to bound the index of U, V, and W.

Claim D.6.2.3. *If Roth's lemma is true for polynomials in $m-1$ or fewer variables, then*

$$\operatorname{Ind}_{\beta_m, r_m}(U) \le k\eta^{2^{m-1}} \quad \text{and} \quad \operatorname{Ind}_{(\beta_1,\dots,\beta_{m-1}, r_1,\dots,r_{m-1})}(V) \le 2k(m-1)\eta^2;$$

and hence the index of W with respect to $(\beta_1,\dots,\beta_m; r_1,\dots,r_m)$ satisfies

$$\begin{aligned}\operatorname{Ind} W &= \operatorname{Ind}_{\beta_m, r_m}(U) + \operatorname{Ind}_{(\beta_1,\dots,\beta_{m-1}, r_1,\dots,r_{m-1})}(V)\\ &\le 2k(m-1)\eta^2 + k\eta^{2^{m-1}}.\end{aligned}$$

PROOF (of Claim D.6.2.3). We want to apply Roth's lemma to V, a polynomial in $m' = m - 1$ variables, with $r'_j = kr_j$ and $\eta' = \eta^2$. We have $\deg_{X_j}(V) \le r'_j$ from (D.6.2.2). We first check condition (6.2-i),

$$\frac{r'_{j+1}}{r'_j} = \frac{r_{j+1}}{r_j} \le \eta^{2^{m-1}} = \eta'^{2^{m'-1}}.$$

Next we check condition (6.2-ii),

$$d'_j h(\beta_j) = kd_j h(\beta_j) \ge k\eta^{-2^{m-1}}(h(P) + 2mr_1) = k\eta'^{-2^{m'-1}}(h(P) + 2mr_1).$$

Now we observe that the inequality $h(U) \le k(h(P) + 2r_1)$ implies that $k(h(P) + 2mr_1) \ge h(U) + 2m'kr_1$, which completes the verification of condition (6.2-ii). By induction we conclude that

$$\operatorname{Ind}_{(r_1,\ldots,r_{m-1})}(V) = k\operatorname{Ind}_{(r'_1,\ldots,r'_{m-1})}(V) \le k(2m'\eta') = 2k(m-1)\eta^2.$$

We now want to apply Roth's lemma in one variable to U with $\eta'' = \eta^{2^{m-1}}$ and $r'' = kr_m$. We have $\deg_{X_m}(U) \le r'_m$, from (D.6.2.2). Note that condition (6.2-i) is empty when $m = 1$, so we only need to check condition (6.2-ii), for which purpose we may use the improved version of Roth's lemma in one variable. Thus

$$\begin{aligned} h(U) + r'' &\le k(h(P) + c_1 r_1) + kr_m \le k(h(P) + 2r_1) \\ &\le \eta^{2^{m-1}} kr_m h(\beta_m) = \eta'' r'' h(\beta_m) \end{aligned}$$

(where $c_1 = \frac{4}{3}\log 2 + \frac{1}{2} \approx 1.4242$ and $c_1 + \eta^{2^{m-1}} \le c_1 + \frac{1}{4} \le 2$). We apply Roth's lemma for a polynomial in one variable (which we already proved) and conclude that

$$\operatorname{Ind}_{(\beta_m,r_m)}(U) = k\operatorname{Ind}_{(\beta_m,r'')}(U) \le k\eta'' = k\eta^{2^{m-1}}.$$

This completes the proof of the claim. □

The next step is to relate the index of W back to the index of P. It is clear from the definition of W that if P vanishes to high order at a point $(\beta_1, \ldots, \beta_m)$, then the same will be true of every entry in the matrix defining W, and so the same will be true for W itself. We need to quantify this observation, as in the following result.

Claim D.6.2.4. *With notation as above, we have*

$$\operatorname{Ind} W \ge \frac{k}{2} \min\left\{\operatorname{Ind} P, (\operatorname{Ind} P)^2\right\} - k\frac{r_m}{r_{m-1}}.$$

PROOF (of Claim D.6.2.4). We begin by estimating the index of a typical entry in the matrix for W with respect to $(\beta_1,\dots,\beta_m;r_1,\dots,r_m)$:

$$\begin{aligned}
&\operatorname{Ind}\left(\Delta_i'\left(\frac{1}{(j-1)!}\frac{\partial^{j-1}}{\partial X_m^{j-1}}P\right)\right)\\
&\quad= \operatorname{Ind}\partial_{i_1,\dots,i_{m-1},j-1}P\\
&\quad\ge \operatorname{Ind}P-\frac{i_1}{r_1}-\cdots-\frac{i_{m-1}}{r_{m-1}}-\frac{j-1}{r_m}\quad\text{from lemma D.3.2(a)}\\
&\quad\ge \operatorname{Ind}P-\frac{i_1+\cdots+i_{m-1}}{r_{m-1}}-\frac{j-1}{r_m}\quad\text{since } r_1\ge r_2\ge\cdots \text{ from (6.2-iii)}\\
&\quad\ge \operatorname{Ind}P-\frac{r_m}{r_{m-1}}-\frac{j-1}{r_m}\\
&\qquad\qquad\text{since } \operatorname{order}(\Delta_i')=i_1+\cdots+i_{m-1}\le i-1\le k-1\le r_m.
\end{aligned}$$

Now each entry in the jth column of the matrix defining W has the form $\partial_{i_1,\dots,i_{m-1},j-1}P$, and W itself looks like

$$W=\sum_{k!\text{ terms}}\begin{pmatrix}\text{product of } k \text{ polynomials, one from}\\ \text{each column of the matrix defining } W\end{pmatrix}.$$

The previous calculation gives a lower bound for the index of each of the entries in the matrix defining W. Hence the index of W with respect to $(\beta_1,\dots,\beta_m;r_1,\dots,r_m)$ satisfies

$$\begin{aligned}
\operatorname{Ind}W&\ge\min_{\substack{k!\text{ terms in}\\ \text{sum for } W}}\left\{\operatorname{Ind}\begin{pmatrix}\text{product of } k \text{ polynomials of the form}\\ \partial_{i_1,\dots,i_{m-1},j-1}P \text{ with } j=1,2,\dots,k\end{pmatrix}\right\}\\
&\qquad\qquad\text{from Lemma D.3.2(b), which says}\\
&\qquad\qquad\text{that } \operatorname{Ind}(\textstyle\sum F_i)\ge\min\{\operatorname{Ind}F_i\}\\
&\ge\sum_{j=1}^{k}\min_{i_1,\dots,i_{m-1}}\operatorname{Ind}\partial_{i_1,\dots,i_{m-1},j-1}P\quad\text{from Lemma D.3.2(c), which}\\
&\qquad\qquad\text{says that } \operatorname{Ind}(\textstyle\prod F_i)=\sum\operatorname{Ind}F_i
\end{aligned}$$

(N.B. It is crucial here to have a sum over $j=1,\dots,k$, rather than just taking k times the minimum index of all entries in the matrix for W. We will indicate below why this is so important.)

Substituting in the lower bound obtained above for $\operatorname{Ind}\partial_{i_1,\dots,i_{m-1},j-1}P$ in those cases where it is positive, we obtain

$$\begin{aligned}
\operatorname{Ind}W&\ge\sum_{j=1}^{k}\max\left\{\operatorname{Ind}P-\frac{r_m}{r_{m-1}}-\frac{j-1}{r_m},0\right\}\\
&\ge\sum_{j=1}^{k}\max\left\{\operatorname{Ind}P-\frac{j-1}{r_m},0\right\}-\frac{kr_m}{r_{m-1}}.
\end{aligned}$$

Combining this with the upper bound for $\operatorname{Ind} W$ given in Claim D.6.2.2 gives the fundamental inequality

$$\sum_{j=1}^{k} \max\left\{\operatorname{Ind} P - \frac{j-1}{r_m}, 0\right\} \leq \operatorname{Ind} W + \frac{kr_m}{r_{m-1}}.$$

It is clear that this inequality says that the index of P cannot be too large. It suffices now to show that,

$$\sum_{j=1}^{k} \left(\operatorname{Ind} P - \frac{j-1}{r_m}\right) \geq \frac{k}{2} \min\left\{\operatorname{Ind} P, (\operatorname{Ind} P)^2\right\}.$$

We consider two cases.

Case 1. $\operatorname{Ind} P \geq \dfrac{k-1}{r_m}$

In this case we obtain

$$\sum_{j=1}^{k} \left(\operatorname{Ind} P - \frac{j-1}{r_m}\right) = k \operatorname{Ind} P - \frac{(k-1)k}{2r_m} \geq \frac{k}{2} \operatorname{Ind} P,$$

where the last inequality follows from the assumption made in the case under consideration, namely that $\operatorname{Ind} P \geq (k-1)/r_m$.

Case 2. $\operatorname{Ind} P \leq \dfrac{k-1}{r_m}$

Let $N = [r_m \operatorname{Ind} P]$, so our assumption implies that $N \leq k-1$. Then our fundamental inequality becomes

$$\begin{aligned}
&\sum_{j=1}^{N+1} \left(\operatorname{Ind} P - \frac{j-1}{r_m}\right) \\
&\quad = (N+1) \operatorname{Ind} P - \frac{N(N+1)}{2r_m} \\
&\quad = (N+1) \left(\operatorname{Ind} P - \frac{[r_m \operatorname{Ind} P]}{2r_m}\right) \qquad \text{from definition of } N \\
&\quad \geq (N+1) \cdot \frac{1}{2} \operatorname{Ind} P \\
&\quad \geq r_m \operatorname{Ind} P \cdot \frac{1}{2} \operatorname{Ind} P \quad \text{from definition of } N \\
&\quad \geq \frac{k}{2} (\operatorname{Ind} P)^2 \quad \text{if we have } k \leq r_m.
\end{aligned}$$

There remains the possibility that $k = r_m + 1$. In this case the quantity we wish to estimate is

$$q(N) := \sum_{j=1}^{N+1} \left(\operatorname{Ind} P - \frac{j-1}{r_m}\right) = (N+1) \operatorname{Ind} P - \frac{N(N+1)}{2(k-1)}.$$

Notice that $q(N)$ is a quadratic function of N. We now observe that

$$(k-1)\operatorname{Ind}P - 1 \le N \le (k-1)\operatorname{Ind}P,$$

while a straightforward computation gives

$$q\big((k-1)\operatorname{Ind}P - 1\big) = q((k-1)\operatorname{Ind}P) = \frac{(k-1)(\operatorname{Ind}P)^2 + \operatorname{Ind}P}{2}.$$

Therefore, since $\operatorname{Ind}P \le 1$ (in the case we are considering), we find that

$$q(N) \ge \frac{(k-1)(\operatorname{Ind}P)^2 + \operatorname{Ind}P}{2} \ge \frac{k(\operatorname{Ind}P)^2}{2},$$

which completes the proof of Case 2 and thus of Claim D.6.2.4. □

We can now easily finish the proof of Roth's lemma using the upper and lower bounds for $\operatorname{Ind}W$ furnished respectively by Claims D.6.2.3 and D.6.2.4. Since $\operatorname{Ind}P \le m$, we may use Claim D.6.2.4 to write

$$\operatorname{Ind}W + \frac{kr_m}{r_{m-1}} \ge \frac{k}{2}\min\left\{\operatorname{Ind}P, (\operatorname{Ind}P)^2\right\} \ge \frac{k(\operatorname{Ind}P)^2}{2m},$$

while Claim D.6.2.3 implies that

$$\begin{aligned}\operatorname{Ind}W + \frac{kr_m}{r_{m-1}} &\le 2k(m-1)\eta^2 + k\eta^{2^{m-1}} + \frac{kr_m}{r_{m-1}} \\ &\le k\left(2(m-1)\eta^2 + 2\eta^{2^{m-1}}\right) \le k(2\eta^2 m).\end{aligned}$$

We deduce that $(\operatorname{Ind}P)^2 \le 4\eta^2 m^2$, and hence $\operatorname{Ind}P \le 2m\eta$. □

Remark. We note that in both Cases 1 and 2, it was essential to consider the sum $\sum (j-1)/r_m = (k-1)k/2r_m$, rather than merely $\sum k/r_m = k^2/r_m$. The point is that no matter how the constants are adjusted, having an extra factor of 2 would destroy the argument in at least one of the cases.

D.7. Completion of the Proof of Roth's Theorem

We have now assembled all of the pieces needed to complete the proof of Roth's Theorem (D.2.1), or more precisely, to prove Theorem D.2.2, which we showed was equivalent to Roth's theorem. For the convenience of the reader, we restate the result we will be proving in this section as Theorem D.7.1, although note that we have used δ in place of ε.

Theorem D.7.1. (= Roth's theorem D.2.2) *Let K be a number field, let $S \subset M_K$ be a finite set of absolute values on K with each absolute value extended in some way to $\bar{K}$. Let $\alpha \in \bar{K}$ and $\delta > 0$ be given. Suppose that*

$$\xi : S \to [0,1] \quad \text{is a function satisfying} \quad \sum_{v \in S} \xi_v = 1.$$

Then there are only finitely many $\beta \in K$ with the property that

$$\|\beta - \alpha\|_v \le \frac{1}{H_K(\beta)^{(2+\delta)\xi_v}} \quad \text{for all } v \in S. \tag{$**$}$$

PROOF. We assume that there are infinitely many solutions to $(**)$ and derive a contradiction. The basic strategy is as follows. We pick a large m and take solutions $\beta_1, \dots, \beta_m$ to $(**)$ satisfying certain conditions (namely m is large, β_1 has large height, β_2 has height much larger than β_1, etc.). We use Proposition D.4.3 to produce a polynomial $P(X_1, \dots, X_m)$ vanishing to a high order at $(\alpha, \dots, \alpha)$, we apply Proposition D.5.1 to show that the index of the polynomial P at $(\beta_1, \dots, \beta_m; r_1, \dots, r_m)$ is greater than $m\varepsilon$, and we use Proposition D.6.2 (Roth's lemma) to show that the index is (strictly) less than $m\varepsilon$. This contradiction will show that $(**)$ has only finitely many solutions.

Since we will need to refer to the various conditions described in Propositions D.4.2, D.5.1, and D.6.2, we list them here in somewhat abbreviated form. The constant $B(\alpha)$ is defined in Proposition D.4.3 and the constant $C(\alpha, \delta)$ in Proposition D.5.1:

(4.2-i) $\quad e^{\varepsilon^2 m/4} > 2[\mathbb{Q}(\alpha) : \mathbb{Q}]$.

(4.2-ii) $\quad \operatorname{Ind}(P) \ge \frac{m}{2}(1 - \varepsilon)$ with respect to $(\alpha, \dots, \alpha; r_1, \dots, r_m)$.

(4.2-iii) $\quad |P| \le B(\alpha)^{r_1 + \cdots + r_m}$.

(5.1-i) $\quad 0 < \varepsilon < \frac{\delta}{22}$.

(5.1-ii) $\quad \|\beta_h - \alpha\|_v \le \frac{1}{H_K(\beta_h)^{(2+\delta)\xi_v}}$.

(5.1-iii) $\quad D := \min_{1 \le h \le m} \{H(\beta_h)^{r_h}\} \le \max_{1 \le h \le m} \{H(\beta_h)^{r_h}\} \le D^{1+\varepsilon}$.

(5.1-iv) $\quad H(\beta_h) \ge C(\alpha, \delta)$ for $1 \le h \le m$.

(6.2-i) $\quad r_{h+1} \le \omega r_h$ for $1 \le h \le m - 1$.

(6.2-ii) $\quad \log |P| + 2mr_1 \le \omega \log D$.

Assume now that there are infinitely many solutions to the inequality $(**)$. Decreasing δ only serves to make the theorem stronger, so we may assume that $0 < \delta < 1$. We are going to choose the quantities

$$\varepsilon, m, \omega, \beta_1, \dots, \beta_m, r_1, \dots, r_m, P(X_1, \dots, X_m)$$

in the listed order as follows:

(1) Choose an ε with $0 < \varepsilon < \delta/22$. Then ε satisfies (5.1-i); note also that $\varepsilon < 1/22 < 1$.

(2) Choose an integer m with $e^{\varepsilon^2 m/4} > 2[\mathbb{Q}(\alpha) : \mathbb{Q}]$. Then (4.2-i) is true. We define $\omega = \omega(m, \varepsilon) = (\varepsilon/4)^{2^{m-1}}$, which implies $2\omega^{2^{-m+1}} = \varepsilon/2 < \varepsilon$.

(3) Since by assumption (**) has infinitely many solutions in K, and since K has only finitely many elements of bounded height, we can find a solution β_1 whose height satisfies

$$H(\beta_1) \geq C(\alpha, \delta) \quad \text{and} \quad \log H(\beta_1) \geq \frac{m(\log B(\alpha) + 2)}{\omega}.$$

(4) We then choose successively $\beta_2, \ldots, \beta_m$ to be solutions to (**) satisfying

$$H_K(\beta_{h+1})^{\omega} \geq H_K(\beta_h)^2 \quad \text{for all } 1 \leq h < m.$$

Notice that since $\omega < 1$, we will have $H_K(\beta_h) \geq H_K(\beta_1)$. In fact, the sequence of $H_K(\beta_h)$'s will be increasing, and hence (5.1-iv) will be satisfied in view of the choice made in (3).

(5) Choose an integer r_1 satisfying $H_K(\beta_1)^{\omega r_1} \geq H_K(\beta_m)^2$.

(6) We want to choose $r_2, \ldots, r_m$ in such a way that all of the $H_K(\beta_h)^{r_h}$'s are approximately equal. So we define $r_1, \ldots, r_m$ to be the integers

$$r_h = \left\lceil \frac{r_1 \log H_K(\beta_1)}{\log H_K(\beta_h)} \right\rceil = \left\lceil \frac{r_1 \log H(\beta_1)}{\log H(\beta_h)} \right\rceil.$$

Here $\lceil t \rceil$ denotes the ceiling of t, that is, the smallest integer that is greater than or equal to t. In order to check conditions (5.1-iii), we compute

$$\begin{aligned}
&r_1 \log H_K(\beta_1) \\
&\quad \leq r_h \log H_K(\beta_h) \quad \text{definition of } r_h \text{ and } \lceil t \rceil \geq t \\
&\quad \leq r_1 \log H_K(\beta_1) + \log H_K(\beta_h) \quad \text{definition of } r_h \text{ and } \lceil t \rceil \leq t + 1 \\
&\quad \leq r_1 \log H_K(\beta_1) + \log H_K(\beta_m) \quad \text{since } H_K(\beta_h)\text{'s increase from (4)} \\
&\quad \leq (1 + \varepsilon) r_1 \log H_K(\beta_1) \quad \text{from the choice of } r_1 \text{ in (5).}
\end{aligned}$$

Exponentiating gives (5.1-iii). Finally, we can verify property (6.2-i) as follows:

$$\begin{aligned}
\frac{r_{h+1}}{r_h} &= \frac{\left\lceil \dfrac{r_1 \log H(\beta_1)}{\log H(\beta_{h+1})} \right\rceil}{\left\lceil \dfrac{r_1 \log H(\beta_1)}{\log H(\beta_h)} \right\rceil} \quad \text{from the choice of the } r_h\text{'s} \\
&\leq \left(\frac{r_1 \log H(\beta_1)}{\log H(\beta_{h+1})} + 1 \right) \Big/ \left(\frac{r_1 \log H(\beta_1)}{\log H(\beta_h)} \right) \\
&= \frac{\log H(\beta_h)}{\log H(\beta_{h+1})} + \frac{\log H(\beta_h)}{r_1 \log H(\beta_1)} \\
&\leq \frac{\omega}{2} + \frac{\omega}{2} = \omega \quad \text{from the choice made in (4) and (5).}
\end{aligned}$$

Hence $r_{h+1} \le \omega r_h$, which verifies condition (6.2-i).

(7) Since m was chosen to verify (4.2-i), we can use Proposition D.4.3 to produce a polynomial $P(X_1, \dots, X_m)$ with $\deg_{X_h} P \le r_h$ satisfying (4.2-ii) and (4.2-iii).

(8) We have verified above that our chosen quantities satisfy the four conditions (5.1-i), (5.1-ii), (5.1-iii), and (5.1-iv). Hence we can apply Proposition D.5.1 to conclude that

$$\operatorname{Ind} P \ge m\varepsilon \qquad \text{with respect to } (\beta_1, \dots, \beta_m; r_1, \dots, r_m).$$

(9) We would like to apply Proposition D.6.2 (Roth's lemma). We have verified condition (6.2-i) with $\eta^{2^{m-1}} = \omega$, so it remains to check condition (6.2-ii). We use the fact that

$$\log D = \min_{1 \le h \le m} \{r_h \log H(\beta_h)\} = r_1 \log H(\beta_1) \quad \text{and} \quad r_1 = \max_h \{r_h\}$$

to compute

$$\begin{aligned} \frac{\log |P| + 2mr_1}{\log D} &\le \frac{(r_1 + \cdots + r_m) \log B(\alpha) + 2mr_1}{\log D} \quad \text{from (4.2-iii)} \\ &\le \frac{m (\log B(\alpha) + 2)}{\log H(\beta_1)} \quad \text{since } r_1 > r_2 > \cdots \\ &\le \omega \quad \text{from the choice of } \beta_1 \text{ in (3).} \end{aligned}$$

This completes the verification of all of the conditions necessary to apply Proposition D.6.2 with $\eta = \omega^{2^{-m+1}} = \varepsilon/4$, so we conclude that

$$\operatorname{Ind} P \le 2m\eta = m\varepsilon/2 \quad \text{with respect to } (\beta_1, \dots, \beta_m; r_1, \dots, r_m).$$

We now observe that the lower and upper bounds for the index of P given in (8) and (9) contradict each other. This completes the proof of Theorem D.2.2 that (**) has only finitely many solutions. Then using the reduction lemma (D.2.2.1), we conclude that Theorem D.2.1 (Roth's theorem) is also true.

□

Remark D.7.2. The proof of Roth's theorem is not effective. This means that for a given α, it does not provide a method that is guaranteed to find all $\beta \in K$ satisfying the inequality

$$\prod_{v \in S} \min\{\|\beta - \alpha\|_v, 1\} \le \frac{1}{H_K(\beta)^{2+\delta}}.$$

Looking at the proof, it is easy to see why this happens. In order to arrive at finiteness of solutions, we first assumed that there is a solution β_1 whose

height is very large. Note that we can specify how large it must be in terms of quantities depending only on K, S, α, and δ. Then we assumed that there is a second solution β_2 whose height is much larger than that of β_1, namely $H_K(\beta_2) \ge H_K(\beta_1)^{2/\omega}$. Note that we cannot determine how large β_2 needs to be until we find a β_1. Similarly, the lower bound for the height of β_3 depends on the height of β_2, and so on. Therein lies the crux of the problem. Suppose, for example, that there are actually no β's satisfying the inequality. In principle, that would be wonderful; but in practice, we would never be able to rigorously demonstrate that there are none! This is the reason that Roth's theorem is ineffective. And until some new method of proof becomes available, it is likely to remain so.

Although Roth's theorem is ineffective, it is possible to give explicit bounds for the *number* of solutions. The reason this can be done is that the β's satisfy a gap principle, which means that their heights grow very rapidly. (See Exercises D.12 and D.13.) Intuitively, this lets one bound the maximum number of solutions between the unknown β_r's without knowing the values of the (possibly nonexistent) β_r's. The first result giving a bound for the number of exceptions to Roth's theorem is due to Davenport and Roth [1], and subsequent improvements have been made by Lewis and Mahler [1], Mignotte [1], Bombieri and van der Poorten [1], Silverman [3], and R. Gross [1]. The following is a typical result. We will not give the proof, but see Exercises D.14 and D.15, as well as the cited references.

Theorem D.7.3. *Let K, S, α, and δ be as in the statement of Roth's theorem. There are effective constants C_1 and C_2, depending only on the numbers $[K(\alpha) : \mathbb{Q}]$ and δ, such that there are at most $4^{\operatorname{card}(S)} C_2$ numbers $\beta \in K$ satisfying the simultaneous inequalities*

$$H_K(\beta) \ge \max\{H_K(\alpha), 2\}^{C_1} \quad \textit{and} \quad \prod_{v \in S} \min\{\|\beta - \alpha\|_v, 1\} \le \frac{1}{H_K(\beta)^{2+\delta}}.$$

D.8. Application: The Unit Equation $U + V = 1$

In this section we begin to reap the rewards for having taken the time to prove Roth's theorem. As our first application, we will show that the two-variable S-unit equation has only finitely many solutions. Then we will use this result to prove that a hyperelliptic curve has only finitely many S-integer points and that a rational function with at least three poles takes on only finitely many S-integral values. In the next section we will use a bit more machinery to extend this last result to algebraic curves of higher genus.

Theorem D.8.1. (Siegel, Mahler) *Let $K/\mathbb{Q}$ be a number field, let $S \subset M_K$ be a finite set of absolute values on K that includes all the archimedean absolute values, and let R_S be the ring of S-integers of K. Then the S-unit equation*

$$U + V = 1$$

has only finitely many solutions in S-units $U, V \in R_S^*$.

Remark D.8.1.1. The original proof of Theorem D.8.1, in the case that the ring R_S is the ring of integers of K, is due to Siegel [1]. Subsequently Mahler [1, 2] gave the generalization to p-adic absolute values. Both Siegel and Mahler used their results as an intermediate step in proving that there are only finitely many integral points on certain curves. The importance of the unit equation as an object of study was pointed out by Lang [1], who also generalized Siegel's and Mahler's finiteness theorems to finitely generated fields.

Remark D.8.1.2. Using the fact that R_S^* is a finitely generated group, we can see why the theorem "should" be true. We are looking at the group variety $\mathbb{G}_m \times \mathbb{G}_m$, and inside this group variety we are looking at the intersection of the finitely generated group $R_S^* \times R_S^*$ and the proper subvariety $\{U+V=1\}$. It is thus unlikely that there should be very many points of the group lying on the subvariety. The theorem asserts that there are only finitely many such points.

To help the reader understand the ideas underlying the proof of Theorem D.8.1, we now briefly sketch the argument. Suppose that there are infinitely many solutions $(U, V) \in R_S^*$ to the S-unit equation $U + V = 1$. We want to use Roth's theorem to derive a contradiction. How can we relate the solutions (U, V) of the unit equation to Diophantine approximation and Roth's theorem? The fact that U and V are S-units means that there is some absolute value $w \in S$ for which $|U|_w$ and $|V|_w$ are large. We then find that

$$\left|\frac{U}{V} - 1\right|_w = \frac{1}{|V|_w}$$

is small, so U/V is a good approximation to 1. Of course, this naive inequality does not contradict Roth's theorem. So we use the finite generation of R_S^* to replace U and V by aX^m and bY^m for some large integer m. Then we show that X/Y is almost as close to $\sqrt[m]{-b/a}$ as U/V is to 1, while the height $H_K(X/Y)$ is approximately $H_K(U/V)^{1/m}$. In other words, by taking m^{th}-roots, we make the height much smaller without affecting the approximating distance too much. Taking m large enough will then contradict Roth's theorem.

PROOF (of Theorem D.8.1). Suppose that there are infinitely many solutions $(U, V) \in R_S^*$ to the S-unit equation $U + V = 1$. We will derive a contradiction. Let $s = \#S$ be the number of absolute values in S, and fix a

"large" integer m. For example, $m = 2s+1$ will suffice. The unit group R_S^* is finitely generated, so the quotient group R_S^*/R_S^{*m} is finite. Fix a set of coset representatives $\mathcal{A}$ for R_S^*/R_S^{*m}. Then every element in R_S^* can be written uniquely as an element of $\mathcal{A}$ multiplied by an mth power, so we can define a map

$$\begin{aligned} \{(U,V) \in R_S^* \times R_S^* \mid U+V=1\} &\longrightarrow \mathcal{A} \times \mathcal{A}, \\ (U,V) &\longmapsto (a,b) \quad \text{with } U/a, V/b \in R_S^{*m}. \end{aligned}$$

By assumption, the set on the left is infinite, while $\mathcal{A}$ is finite. The pigeonhole principle says that we can choose some $a,b \in \mathcal{A}$ such that there are infinitely many (U,V)'s that map to (a,b). Writing $U/a = X^m$ and $V/b = Y^m$, this shows that there exists some $a,b \in \mathcal{A}$ such that the equation

$$aX^m + bY^m = 1$$

has infinitely many solutions $X,Y \in R_S^*$. We will derive a contradiction by showing that X/Y is too good an approximation to $\sqrt[m]{-b/a}$.

The set S contains only finitely many absolute values, so applying the pigeonhole principle again, we can find an absolute value $w \in S$ such that the equation $aX^m + bY^m = 1$ has infinitely many solutions (X,Y) with

$$\|Y\|_w = \max\{\|Y\|_v \mid v \in S\}$$

(i.e., the pigeons are the solutions (X,Y), the pigeonholes are the elements of S, and we assign a pigeon (X,Y) to a pigeonhole by choosing the v in S that maximizes $\|Y\|_v$).

To ease notation, we fix an mth-root $\alpha = \sqrt[m]{-b/a}$. Then

$$\frac{1}{aY^m} = \frac{X^m}{Y^m} + \frac{b}{a} = \frac{X^m}{Y^m} - \alpha^m = \prod_{\zeta \in \mu_m} \left(\frac{X}{Y} - \zeta\alpha \right),$$

where the product is over all m^{th}-roots of unity. Clearly, if Y has large absolute value, then at least one of the factors $X/Y - \zeta\alpha$ must be small. We claim that only one of them can be very small.

To see this, let $\zeta, \zeta' \in \mu_m$ be distinct m^{th}-roots of unity and use the triangle inequality to compute

$$\left| \frac{X}{Y} - \zeta\alpha \right|_w + \left| \frac{X}{Y} - \zeta'\alpha \right|_w \ge |\zeta'\alpha - \zeta\alpha|_w \ge C_1.$$

Here the constant $C_1 = C_1(K,S,m)$ can be chosen in terms of K, S, and m, independently of X and Y. (In principle, C_1 also depends on a and b. But a and b are chosen from a set of coset representatives for R_S^*/R_S^{*m}, so they may be determined by S.) It follows that

$$\frac{1}{|aY^m|_w} = \prod_{\zeta \in \mu_m} \left| \frac{X}{Y} - \zeta\alpha \right|_w \ge \left(\min_{\zeta \in \mu_m} \left| \frac{X}{Y} - \zeta\alpha \right|_w \right) \cdot \left(\frac{C_1}{2} \right)^{m-1},$$

since all but one of the terms in the product must be greater than $C_1/2$. Hence

$$\frac{1}{\|Y\|_w^m} \geq C_2 \min_{\zeta\in\mu_m} \left\| \frac{X}{Y} - \zeta\alpha \right\|_w ,$$

where the constant $C_2 = C_2(K, S, m)$ depends on K, S, and m.

We invoke the pigeonhole principle one more time, again with pigeons (X, Y), but this time the pigeonholes are the mth-roots of unity. We assign a pigeon (X, Y) to the pigeonhole $\zeta \in \mu_m$ that minimizes $\|(X/Y) - \zeta\alpha\|_w$. Some pigeonhole, call it ξ, will have infinitely many pigeons. Hence there are infinitely many solutions $X, Y \in R_S^*$ to the equation $aX^m + bY^m = 1$ satisfying

$$\frac{1}{\|Y\|_w^m} \geq C_2 \left\| \frac{X}{Y} - \xi\alpha \right\|_w .$$

This shows that X/Y is a good approximation to $\xi\alpha$. In order to apply Roth's theorem, we must relate $\|Y\|_w$ to the height of X/Y.

The absolute value w was chosen to maximize $\|Y\|_v$. Since we also have $\|Y\|_v = 1$ for all $v \notin S$, it follows that

$$\|Y\|_w = \max_{v\in S} \|Y\|_v \geq \Big(\prod_{v\in S} \|Y\|_v\Big)^{1/s} = \Big(\prod_{v\in M_K} \|Y\|_v\Big)^{1/s} = H_K(Y)^{1/s}.$$

Further, using elementary properties of height functions, namely

$$H(x+y) \leq 2H(x)H(y) \quad \text{and} \quad H(xy) \leq H(x)H(y)$$

(see Exercise B.20), and the fact that (X, Y) is a solution of $aX^m + bY^m = 1$, we compute

$$\begin{aligned} H_K\left(\frac{X^m}{Y^m}\right) &= H_K\left(\frac{1}{aY^m} - \frac{b}{a}\right) \leq 2^{[K:\mathbb{Q}]} H_K\left(\frac{1}{aY^m}\right) H_K\left(\frac{b}{a}\right) \\ &\leq 2^{[K:\mathbb{Q}]} H_K\left(\frac{1}{Y^m}\right) H_K\left(\frac{1}{a}\right) H_K\left(\frac{b}{a}\right). \end{aligned}$$

Taking m^{th} roots and using the fact that $H_K(T^m) = H_K(T)^m$, we find that there is a constant $C_3 = C_3(K, S, m)$ such that

$$H_K(X/Y) \leq C_3 H_K(1/Y) = C_3 H_K(Y).$$

Combining this with the bound for $H_K(Y)$ given above, we obtain

$$\|Y\|_w \geq C_4 H_K(X/Y)^{1/s},$$

where $C_4 = C_4(K, S, m) = C_3^{-1/s}$.

We showed above that $1/\|Y\|_w^m \ge C_2\|X/Y - \xi\alpha\|_w$, so using the estimate for $\|Y\|_w$, we get

$$\frac{C_5}{H_K(X/Y)^{m/s}} \ge \left\| \frac{X}{Y} - \xi\alpha \right\|_w$$

with $C_5 = C_5(K,S,m) = 1/C_2C_4^m$. Our assumption is that there are infinitely many $X, Y \in R_S^*$ satisfying this inequality. We now recall that we chose $m = 2s+1$, so Roth's Theorem (D.2.1) tells us that this last inequality has only finitely many solutions in K. This contradiction concludes the proof of Theorem D.8.1. □

Theorem D.8.1 says that the equation $U + V = 1$ has only finitely many solutions $U, V \in R_S^*$. The following quantitative version, whose proof is beyond the scope of this book, gives a very strong upper bound for the number of solutions. Notice that the bound depends only on the degree of K over $\mathbb{Q}$ and on the number of places in S. It is independent of A, B, and the particular places in S.

Theorem D.8.2. (Evertse [1]) *Let $K/\mathbb{Q}$ be a number field, let $S \subset M_K$ be a finite set of absolute values on K that includes all archimedean absolute values, and let R_S be the ring of S-integers of K. Then for any $A, B \in K^*$, the S-unit equation*

$$AU + BV = 1$$

has at most $3 \cdot 7^{[K:\mathbb{Q}]+2\#S}$ solutions in S-units $U, V \in R_S^*$.

We will now give Siegel's proof that a hyperelliptic curve has only finitely many integer points. Although this result will be superseded by Theorem D.9.1 in the next section, the proof by reduction to the S-unit equation is instructive. Further, the effective solution of the S-unit equation using linear forms in logarithms leads to effective bounds for the size of integer points on hyperelliptic curves. The reader should note how the proof uses the two fundamental finiteness theorems of algebraic number theory, namely the finiteness of the ideal class group and the finite generation of the unit group (whose proofs were given in Part C, Section 3).

Theorem D.8.3. (Siegel) *Let $K/\mathbb{Q}$ be a number field, let $S \subset M_K$ be a finite set of absolute values on K that includes all the archimedean absolute values, and let R_S be the ring of S-integers of K. Let $f(X) \in K[X]$ be a polynomial of degree at least 3 with distinct roots (in $\bar{K}$). Then the equation*

$$Y^2 = f(X) \quad \textit{has only finitely many solutions } X, Y \in R_S.$$

PROOF. Note that the statement of the theorem becomes stronger if we replace K by a finite extension or replace S by a larger set of absolute

values. So we begin by taking an extension of K over which $f(X)$ factors as

$$f(X) = a(X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_n) \quad \text{with } \alpha_1, \ldots, \alpha_n \in K.$$

By assumption, $n \geq 3$ and the α_i's are all distinct.

Next we increase the size of S so that the following three conditions are true.

(i) $a \in R_S^*$ and $\alpha_1, \ldots, \alpha_n \in R_S$.
(ii) $\alpha_i - \alpha_j \in R_S^*$ for all $i \neq j$.
(iii) R_S is a principal ideal domain.

Clearly, conditions (i) and (ii) only require us to add a finite number of primes to S. The same is true of condition (iii), since in any case the ideal class group of R_S is finite, and it suffices to add to S one prime ideal from each ideal class. (If you do not want to use Dirichlet's theorem on primes in arithmetic progressions, just take one ideal $\mathfrak{a}$ from each ideal class and add to S all of the prime ideals dividing $\mathfrak{a}$; see, for example, Lemma C.3.7.)

For later use, we will also define a field L/K by

$$L = K\left(\sqrt{u} \,\middle|\, u \in R_S^*\right).$$

It is vitally important to observe that L is a finite extension of K. This is true because L is generated by the square roots of coset representatives for R_S^*/R_S^{*2}, and we know that R_S^*/R_S^{*2} is finite by Dirichlet's unit theorem. We let $T \subset M_L$ be the set of places of L lying over S, and we write R_T for the ring of T-integers in L.

We now begin the proof of Theorem D.8.3. Suppose that $X, Y \in R_S$ is a solution to the equation $Y^2 = f(X)$. We observe that if $\mathfrak{p}$ is a prime ideal in R_S and if $\mathfrak{p}$ divides $X - \alpha_i$, then for any $j \neq i$ we have

$$X - \alpha_j = (X - \alpha_i) + (\alpha_i - \alpha_j) \equiv \alpha_i - \alpha_j \not\equiv 0 \pmod{\mathfrak{p}}.$$

(The fact that $\alpha_i - \alpha_j \not\equiv 0 \pmod{\mathfrak{p}}$ follows from property (ii) above.) Hence at most one of the $X - \alpha_i$'s can be divisible by $\mathfrak{p}$.

On the other hand, the product of the $X - \alpha_i$'s is equal to $a^{-1}Y^2$, and a is a unit, so the highest power of $\mathfrak{p}$ dividing the product must have even exponent. It follows that every prime dividing the ideal $(X - \alpha_i)R_S$ must divide it to an even power. Therefore, there are ideals $\mathfrak{a}_i \subset R_S$ such that

$$(X - \alpha_i)R_S = \mathfrak{a}_i^2 \quad \text{for all } 1 \leq i \leq n.$$

Property (iii) says that R_S is a principal ideal domain, so $\mathfrak{a}_i = Z_i R_S$ for some $Z_i \in R_S$. Hence there are units $U_i \in R_S^*$ such that $X - \alpha_i = U_i Z_i^2$. To recapitulate, we have shown that

$$X - \alpha_i = U_i Z_i^2 \quad \text{for some } U_i \in R_S^* \text{ and } Z_i \in R_S.$$

It is now apparent why we defined the field L, because in L the unit U_i becomes a square, say $U_i = V_i^2$. Thus we have

$$X - \alpha_i = V_i^2 Z_i^2 = W_i^2 \quad \text{for some } W_i \in R_T,$$

where we have set $W_i = V_i Z_i$. Taking the difference of two of these equations then gives

$$\alpha_j - \alpha_i = W_i^2 - W_j^2 = (W_i - W_j)(W_i + W_j).$$

But condition (ii) says that $\alpha_j - \alpha_i$ is a unit in R_S, so we conclude that both $W_i - W_j$ and $W_i + W_j$ are units,

$$W_i - W_j, W_i + W_j \in R_T^* \quad \text{for all } i \neq j.$$

We now use the fact that $f(X)$ has degree at least three to write down the identity

$$\frac{W_1 - W_2}{W_1 - W_3} + \frac{W_2 - W_3}{W_1 - W_2} = 1.$$

(This is sometimes called *Siegel's identity.*) Each of the terms on the left-hand side is a unit in R_T, so Theorem D.8.1 tells us that each can take on only finitely many values. Similarly, the identity

$$\frac{W_1 + W_2}{W_1 - W_3} + \frac{W_3 + W_2}{W_3 - W_1} = 1$$

and Theorem D.8.1 tell us that each of the terms in this equation can take on only finitely many values. It follows that there are only finitely many possible values for the quantity

$$\frac{W_1 - W_2}{W_1 - W_3} \cdot \frac{W_1 + W_2}{W_1 - W_3} = \frac{W_1^2 - W_2^2}{(W_1 - W_3)^2} = \frac{\alpha_2 - \alpha_1}{(W_1 - W_3)^2}.$$

Therefore, there are only finitely many possible values for $W_1 - W_3$, so finitely many for

$$\frac{1}{2}\left((W_1 - W_3) + \frac{\alpha_3 - \alpha_1}{W_1 - W_3}\right) = \frac{1}{2}((W_1 - W_3) + (W_1 + W_3)) = W_1,$$

and so finitely many for $\alpha_1 + W_1^2 = X$. Finally, for a given X, there are at most two possible values for Y, which completes the proof that $Y^2 = f(X)$ has only finitely many solutions $X, Y \in R_S$. □

Using a similar argument, we investigate when a function on $\mathbb{P}^1$ can assume infinitely many integral values. The reader should compare this result with Theorem D.9.1 in the next section, which deals with curves of genus greater than zero.

Theorem D.8.4. (Siegel) *Let $K/\mathbb{Q}$ be a number field, let $S \subset M_K$ be a finite set of absolute values on K that includes all the archimedean absolute values, and let R_S be the ring of S-integers of K. Let C/K be a curve of genus zero, and let $\phi \in K(C)$ be a rational function on C with at least three distinct poles (in $C(\bar{K})$). Then there are only finitely many rational points $T \in C(K)$ satisfying $\phi(T) \in R_S$.*

PROOF. If $C(K) = \emptyset$, there is nothing to prove. So we may take $C = \mathbb{P}^1$ and write $\phi = f(x,y)/g(x,y)$ using homogeneous polynomials $f, g \in K[x,y]$ of the same degree with no common roots (in $\mathbb{P}^1(\bar{K})$). Taking a finite extension of K and adding finitely many primes to S, we may assume that the following conditions are true.

(i) f and g factor completely in K,

$$f = a(x-\alpha_1 y)^{d_1} \cdots (x-\alpha_m y)^{d_m}, \quad g = b(x-\beta_1 y)^{e_1} \cdots (x-\beta_n y)^{e_n}.$$

(If ϕ has a zero or pole at $[1,0]$, then one of f or g may also have a y^d factor.)

(ii) $a, b \in R_S^*$ and $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \in R_S$.

(iii) $\alpha_i - \beta_j \in R_S^*$ for all $1 \le i \le m$ and $1 \le j \le n$.

(iv) R_S is a principal ideal domain.

Now suppose that $T \in \mathbb{P}^1(K)$ has the property that $\phi(T) \in R_S$. We write $T = [X,Y]$ with $\gcd(X,Y) = 1$, which we can do because R_S is a principal ideal domain. Note that for any i, j we have

$$(X - \alpha_i Y) - (X - \beta_j Y) = (\alpha_i - \beta_j)Y$$

and

$$-\beta_j(X - \alpha_i Y) + \alpha_i(X - \beta_j Y) = (\alpha_i - \beta_j)X.$$

We know from (iii) that $\alpha_i - \beta_j$ is a unit, and we have chosen X and Y to be relatively prime. It follows that $X - \alpha_i Y$ and $X - \beta_j Y$ are relatively prime. But $f(X,Y) = \prod(X - \alpha_i Y)^{d_i}$ and $g(X,Y) = \prod(X - \beta_j Y)^{e_j}$, so $f(X,Y)$ and $g(X,Y)$ are relatively prime in R_S.

On the other hand, we are assuming that $\phi(T) = f(X,Y)/g(X,Y)$ is in R_S, which means that $g(X,Y)$ divides $f(X,Y)$. It follows that $g(X,Y)$ is a unit, $g(X,Y) \in R_S^*$, and so each of the $X - \beta_j Y$'s is in R_S^*. We are further assuming that g has at least three distinct roots, so we can consider Siegel's identity

$$\frac{\beta_2 - \beta_3}{\beta_2 - \beta_1} \cdot \frac{X - \beta_1 Y}{X - \beta_3 Y} - \frac{\beta_3 - \beta_1}{\beta_2 - \beta_1} \cdot \frac{X - \beta_2 Y}{X - \beta_3 Y} = 1.$$

Both terms on the left-hand side are units, so Theorem D.8.1 tells us that they can assume only finitely many values. Finally, if we fix a value for $(X - \beta_1 Y)/(X - \beta_3 Y) = \gamma$, we have $(1-\gamma)X = (\beta_1 - \gamma\beta_3)Y$, so we get only one point

$$T = [X,Y] = [\beta_1 - \gamma\beta_3, 1 - \gamma].$$

(Note that this is a well-defined point, since $\beta_1 \neq \beta_3$.) Hence there are only finitely many $T \in \mathbb{P}^1(K)$ with $\phi(T) \in R_S$. □

D.9. Application: Integer Points on Curves

In this section we will give Siegel's proof that an affine curve of genus at least 1 has only finitely many integer points. We begin by setting the following notation, which will be in effect throughout this section.

$K/\mathbb{Q}$	a number field.
R_S	the ring of S-integers in K for a finite set of places S.
s	$= \#S$, the number of places in S.
C/K	a smooth projective curve of genus g defined over K.
f	a nonconstant function $f \in K(C)$.

Siegel's Theorem D.9.1. *Assume that C has genus $g \geq 1$. Then the set*

$$\{P \in C(K) \mid f(P) \in R_S\}$$

is finite.

Before beginning the proof of Siegel's theorem, we make two quasi-historical remarks.

Remark D.9.2.1. Our proof of Siegel's theorem (D.9.1) will use Roth's theorem. In the 1920s, when he proved his theorem, Siegel had available only the weaker result that an algebraic number α of degree d has approximation exponent $\tau(\alpha) \leq 2\sqrt{d}$, so his original proof was somewhat more complicated. Further, he considered only archimedean absolute values, so his ring R_S was the ring of integers of K.

Remark D.9.2.2. Theorem D.9.1 together with Theorem D.8.4 can be reformulated as follows. Let U be an affine curve of genus g with s points at infinity. In other words, if $\tilde{U}$ is the normalization of U and C a smooth projective curve that is birational to U, then C has genus g and $s = \operatorname{card}(C \smallsetminus \tilde{U})$. Then Theorems D.9.1 and D.8.4 say that

$$2g - 2 + s > 0 \quad \Longrightarrow \quad U(R_{k,S}) \text{ is finite.}$$

Notice that the quantity $2 - 2g - s$ can be viewed as the Euler–Poincaré characteristic $\chi(U)$ of U. Thus the negativity of the very coarse geometric invariant $\chi(U)$ implies a deep arithmetic finiteness property of U.

If $g \geq 2$, then Siegel's theorem is in some sense superseded by Faltings' theorem (né the Mordell conjecture), which asserts that the set $C(K)$ itself

is finite. In fact, Siegel's theorem for curves of genus 0 (with at least 3 points at infinity) or genus 1 (with at least 1 point at infinity) may also be deduced from Faltings' theorem (see Exercise E.11). However, the proof of Siegel's theorem is an instructive lesson in the use of geometry to prove arithmetic results, so it well warrants inclusion here.

A curve of genus 1 is an abelian variety of dimension 1, so Siegel's theorem says that an affine piece of an abelian variety of dimension 1 has only finitely many S-integral points. Lang conjectured that the same should be true for abelian varieties of arbitrary dimension, and Faltings used an adaptation of Vojta's method to prove Lang's conjecture. See Part F and especially Section F.5.3 for further comments.

We begin our proof of Siegel's theorem (D.9.1) with a version of Roth's theorem for curves.

Proposition D.9.3. *With notation as described at the beginning of this section, we let e be the maximum order of the zeros of f, we fix a constant $\varepsilon > 0$, and we choose a function $t \in K(C)$ that is defined and unramified at all zeros and poles of f. Then there exists a constant $c = c(f, t, C, \varepsilon, S) > 0$ such that*

$$\prod_{v \in S} \min\{\|f(P)\|_v, 1\} \geq \frac{c}{H_K\big(t(P)\big)^{(2+\varepsilon)se}} \qquad \text{for all } P \in C(K).$$

PROOF. Write the divisor of f as

$$\operatorname{div}(f) = e_1(Q_1) + e_2(Q_2) + \cdots + e_r(Q_r) - E$$

for some effective divisor $E > 0$. Notice that $Q_1, \ldots, Q_r$ are the distinct zeros of f and that $e = \max\{e_i\}$.

Suppose that the proposition is false. This means that there is a sequence of points $P_1, P_2, \ldots \in C(K)$ such that

$$\lim_{i \to \infty} H_K\big(t(P_i)\big)^{(2+\varepsilon)se} \prod_{v \in S} \min\{\|f(P_i)\|_v, 1\} = 0.$$

We also observe that

$$\prod_{v \in S} \min\{\|f(P_i)\|_v, 1\} \geq \Big(\min_{v \in S}\{\|f(P_i)\|_v, 1\}\Big)^s,$$

so substituting this in and taking s^{th}-roots gives

$$\lim_{i \to \infty} H_K\big(t(P_i)\big)^{(2+\varepsilon)e} \min_{v \in S}\{\|f(P_i)\|_v, 1\} = 0.$$

The height $H_K\big(t(P_i)\big)$ goes to infinity, since $C(K)$ has only finitely many points of bounded height. Hence we can find an absolute value $w \in S$

and a subsequence of the P_i's (which by abuse of notation we again denote by $P_1, P_2, \ldots$) such that

$$\lim_{i\to\infty} H_K\big(t(P_i)\big)^{(2+\varepsilon)e}\big\|f(P_i)\big\|_w = 0.$$

In particular, $\big\|f(P_i)\big\|_w \to 0$, so eventually each P_i must be close to one of the zeros of f in the w-adic topology. Taking a subsequence of the P_i's, we may assume that

$$P_i \underset{i\to\infty}{\longrightarrow} Q \qquad \text{in the } w\text{-adic topology}$$

for some fixed zero Q_j of f.

The function $t - t(Q_j)$ is a uniformizer at Q_j, and f vanishes to order e_j at Q_j, so the function

$$\big(t - t(Q_j)\big)^{-e_j} f$$

has no zero or pole at Q_j. This means that it is w-adically bounded in a sufficiently small w-adic neighborhood of Q_j. In particular, there are constants $c_1, c_2 > 0$ such that

$$c_1 \le \Big\|\big(t(P_i) - t(Q_j)\big)^{-e_j} f(P_i)\Big\|_w \le c_2 \qquad \text{for all sufficiently large } i.$$

It follows from above that

$$\lim_{i\to\infty} H_K\big(t(P_i)\big)^{(2+\varepsilon)e}\Big\|\big(t(P_i) - t(Q_j)\big)^{e_j}\Big\|_w = 0.$$

But e is the largest of the e_j's, so we find that

$$\lim_{i\to\infty} H_K\big(t(P_i)\big)^{(2+\varepsilon)}\big\|t(P_i) - t(Q_j)\big\|_w = 0.$$

This says that the rational numbers $t(P_1), t(P_2), \ldots \in K$ closely approximate the algebraic number $t(Q_j) \in \bar{K}$. In fact, they approximate so closely that they contradict Roth's theorem (D.2.1). This contradiction completes the proof of Proposition D.9.3. □

The next step is to show that the exponent $se(2+\varepsilon)$ in Proposition D.9.3 can be replaced by any positive exponent, provided that we assume that C has positive genus.

Proposition D.9.4. *With notation as described at the beginning of this section, choose a function $t \in K(C)$ that is defined and unramified at all zeros and poles of f, and let $\rho > 0$ be a positive constant. Assume that C has genus $g \ge 1$. Then there exists a constant $c = c(f, t, C, \rho, S) > 0$ such that*

$$\prod_{v\in S} \min\big\{\|f(P)\|_v, 1\big\} \ge \frac{c}{H_K\big(t(P)\big)^\rho} \qquad \text{for all } P \in C(K).$$

PROOF. We begin with a brief sketch of the ideas underlying the proof. We know from (D.9.3) that the result is true for some exponent, specifically $\rho = e(f)s(2+\varepsilon)$, where $e(f)$ is the largest order of vanishing of f. The idea of the proof is to find a covering $\phi : C' \to C$ such that there are rational points $P' \in C'(K)$ lifting the rational points $P \in C(K)$. Then

$$\|f(\phi(P'))\|_v = \|f(P)\|_v.$$

On the other hand, for an appropriate function $t' \in K(C')$ we will find that

$$H_K(t'(P')) \approx H_K(t(P))^{1/\deg(\phi)}.$$

If we apply (D.9.3) to C', $f \circ \phi$, and t', we find that

$$\prod_{v \in S} \min\{\|f \circ \phi(P')\|_v, 1\} \geq \frac{c}{H_K(t'(P'))^{e(f\circ\phi)s(2+\varepsilon)}} \quad \text{for all } P' \in C'(K),$$

so in terms of C we obtain

$$\prod_{v \in S} \min\{\|f(P)\|_v, 1\} \geq \frac{c}{H_K(t(P))^{e(f\circ\phi)s(2+\varepsilon)/\deg(\phi)}} \quad \text{for all } P' \in C'(K).$$

Now by taking the degree of ϕ very large, we can make the exponent as small as we like, provided that the orders of the zeros of $f \circ \phi$ are not much greater than the orders of the zeros of f. A priori, it might happen that ϕ is totally ramified at some zero Q of f, in which case

$$\operatorname{ord}_{Q'}(f \circ \phi) = (\deg \phi) \operatorname{ord}_Q(f) \qquad (\text{where } \phi(Q') = Q).$$

This would vitiate the argument, so we have to prevent it from happening. We will use the multiplication-by-m map on the Jacobian of C to find ϕ's that are everywhere unramified, thereby ensuring that $f \circ \phi$ does not vanish to higher order than f.

We are now ready to begin the proof of Proposition D.9.4. We will suppose that the proposition is false and derive a contradiction. So we suppose that there is a sequence of points $P_0, P_1, P_2, \ldots \in C(K)$ such that

$$H_K(t(P_i))^\rho \prod_{v \in S} \min\{\|f(P_i)\|_v, 1\} \leq c \qquad \text{for all } i = 0, 1, 2, \ldots.$$

We fix an embedding

$$j : C \hookrightarrow J = \operatorname{Jac}(C)$$

defined over K. For example, we could use $j(P) = \operatorname{Cl}(P - P_0)$. Using this embedding, we will treat C as a subvariety of J. We also fix a large

integer m that will be specified later. (In fact, any integer $m > 1 + 6e/\rho$ will suffice.)

The weak Mordell–Weil theorem (C.0.2) says that the quotient group $J(K)/mJ(K)$ is finite, so replacing $\{P_i\}_{i\geq 1}$ with a subsequence, we may assume that every P_i has the same image in $J(K)/mJ(K)$, say

$$P_i = mP_i' + R \qquad \text{for all } i = 1, 2, \ldots.$$

Here $R \in J(K)$ is some fixed rational point on J.

Consider the map

$$\Phi : J \longrightarrow J, \qquad x \longmapsto mx + R.$$

Let C' be the curve $C' = \Phi^{-1}(C)$. Notice that the points $P_1', P_2', \ldots$ are in $C'(K)$. Further, the map Φ is unramified, so its restriction to C',

$$\phi : C' \longrightarrow C,$$

is also unramified. By construction, we have a commutative square

$$\begin{array}{ccc} C' & \overset{j'}{\hookrightarrow} & J \\ \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \Phi} \\ C & \overset{j}{\hookrightarrow} & J \end{array}$$

where the vertical maps are unramified and j' is the natural inclusion of C' in J. In particular, and this is the crucial attribute of this construction, the fact that ϕ is unramified implies that

$$\operatorname{ord}_{P'}(f \circ \phi) = \operatorname{ord}_{\phi(P')}(f) \qquad \text{for all points } P' \in C'.$$

Hence $e(f \circ \phi) = e(f)$, where as usual we are writing $e(\cdot)$ for the highest order of vanishing of a function.

Let $D \in \operatorname{Div}(J)$ be a very ample symmetric divisor on J. Then j^*D is very ample on C, and j'^*D is very ample on C', so we can choose functions $t \in K(C)$ and $t' \in K(C')$ associated to j^*D and j'^*D, respectively. This means that the map $t : C \to \mathbb{P}^1$ satisfies $t^*(\infty) \sim j^*D$, and similarly for t', so we obtain the height relations

$$h\big(t(P)\big) = h_{C,j^*D}(P) + O(1) \quad \text{for all } P \in C(\bar{K}),$$
$$h\big(t'(P')\big) = h_{C',j'^*D}(P') + O(1) \quad \text{for all } P' \in C'(\bar{K}).$$

Next we do a height computation on J to compare $h_{J,D}\circ j$ with $h_{J,D}\circ j'$. For all $P' \in C'(\bar{K})$,

$$\begin{aligned}
h\big(t(\phi(P'))\big) &= h_{C,j^*D}\big(\phi(P')\big) + O(1) && \text{from above}\\
&= h_{J,D}\big(j(\phi(P'))\big) + O(1) && \text{functoriality (B.3.2b)}\\
&= \hat{h}_{J,D}\big(j(\phi(P'))\big) + O(1) && \text{from (B.5.1a)}\\
&= \hat{h}_{J,D}\big(\Phi(j'(P'))\big) + O(1) && \text{since } j\circ\phi = \Phi\circ j'\\
&= \hat{h}_{J,D}\big([m](j'(P')) + R\big) + O(1) && \text{definition of } \Phi\\
&\ge \frac{1}{2}\hat{h}_{J,D}\big([m](j'(P'))\big) - \hat{h}_{J,D}(R) + O(1) && \text{see below}^\dagger\\
&= \frac{m^2}{2}\hat{h}_{J,D}\big(j'(P')\big) + O(1) && \text{from (B.5.1b)}\\
& && \text{(the } O(1) \text{ depends on } R)\\
&= \frac{m^2}{2}h_{J,D}\big(j'(P')\big) + O(m^2) && \text{from (B.5.1a) again}\\
&= \frac{m^2}{2}h_{C,j'^*D}(P') + O(m^2) && \text{from (B.3.2b) again}\\
&= \frac{m^2}{2}h\big(t'(P')\big) + O(m^2) && \text{from above.}
\end{aligned}$$

The inequality marked with a † is a special case of an elementary inequality that says that for any positive definite quadratic form ξ and any real number $t > 0$,

$$\begin{aligned}
\xi(x+y) &= (1-t^{-2})\xi(x) - (t^2-1)\xi(y) + \xi(t^{-1}x + ty)\\
&\ge (1-t^{-2})\xi(x) - (t^2-1)\xi(y).
\end{aligned}$$

We have merely used this estimate with $t = \sqrt{2}$ and $\xi = \hat{h}_{J,D}$.

Exponentiating the above height inequality yields

$$H_K\big(t(\phi(P')\big) \ge \kappa_m H_K\big(t'(P')\big)^{m^2/2} \qquad \text{for all } P' \in C'(\bar{K}).$$

We have written the constant as κ_m to emphasize that it depends on the choice of m. So at some point we will have to fix a value for m that is independent of the sequence of points $P_1', P_2', \ldots \in C'(K)$.

Combining all of the estimates derived above, we compute

$$\begin{aligned}
c &\ge H_K\big(t(P_i)\big)^\rho \prod_{v\in S}\min\big\{\|f(P_i)\|_v, 1\big\} && \text{by assumption}\\
&= H_K\big(t(\phi(P_i'))\big)^\rho \prod_{v\in S}\min\big\{\|f(\phi(P_i'))\|_v, 1\big\} && \text{since } P_i = \phi(P_i')\\
&\ge \kappa_m H_K\big(t'(P_i')\big)^{\rho m^2/2} \prod_{v\in S}\min\big\{\|f(\phi(P_i'))\|_v, 1\big\} && \text{from above}\\
&\ge \kappa_m H_K\big(t'(P_i')\big)^{\rho m^2/2}\cdot c' H_K\big(t'(P_i')\big)^{-(2+\varepsilon)se(f\circ\phi)}\\
& \qquad \text{from (D.9.3) applied to } C',\ f\circ\phi, \text{ and } t'.
\end{aligned}$$

Now we use the crucial fact that ϕ is unramified, so $e(f \circ \phi) = e(f)$ is *independent of* m. This means that we have the inequality

$$c'' \geq H_K\big(t'(P_i')\big)^{\rho m^2/2-(2+\varepsilon)se(f)} \qquad \text{for all } i = 1, 2, \ldots.$$

The constant c'' depends on m, but it is independent of i. On the other hand, the height $H_K\big(t'(P_i')\big)$ goes to infinity as we let $i \to \infty$, so the exponent cannot be positive. Hence

$$\rho m^2/2 \leq (2+\varepsilon)se(f).$$

This holds for every $m \geq 1$, and the right-hand side is independent of m, since we can take $\varepsilon = 1$, for example. Therefore, $\rho \leq 0$. This contradicts our original choice of $\rho > 0$, which completes the proof of Proposition D.9.4. □

With Proposition D.9.4 at our disposal, the proof of Siegel's Theorem D.9.1 is easy.

PROOF (of Siegel's theorem (D.9.1)). We assume that the set

$$\{P \in C(K) \mid f(P) \in R_S\}$$

is infinite and derive a contradiction. Fix a function $t \in K(C)$ that is defined and unramified at all zeros and poles of f, and let $\rho = \deg f/2 \deg t$. (Actually, any ρ satisfying $0 < \rho < \deg f/\deg t$ will do.) Our first step is to apply Proposition D.9.4 to the function $1/f$. We find that there exists a constant $c_1 > 0$ such that

$$\prod_{v \in S} \min\{\|(1/f)(P)\|_v, 1\} \geq \frac{c_1}{H_K\big(t(P)\big)^\rho} \qquad \text{for all } P \in C(K).$$

A small amount of algebra then gives

$$H_K\big(t(P)\big)^\rho \geq c_1 \prod_{v \in S} \max\{\|f(P)\|_v, 1\} \qquad \text{for all } P \in C(K).$$

Next we observe that if $f(P) \in R_S$, then $\big\|f(P)\big\|_v \leq 1$ for all $v \notin S$, so the height of $f(P)$ is given by

$$H_K\big(f(P)\big) = \prod_{v \in M_K} \max\{\|f(P)\|_v, 1\} = \prod_{v \in S} \max\{\|f(P)\|_v, 1\}.$$

Combining this with the previous inequality gives

$$H_K\big(t(P)\big)^\rho \geq c_1 H_K\big(f(P)\big) \qquad \text{for all } P \in C(K) \text{ with } f(P) \in R_S.$$

Now we take logarithms and divide by the degree $[K:\mathbb{Q}]$ to obtain an estimate in terms of logarithmic heights,

$$\rho h\big(t(P)\big) \ge h\big(f(P)\big) - c_2 \qquad \text{for all } P \in C(K) \text{ with } f(P) \in R_S.$$

Recall that we chose ρ to equal $\deg f / 2 \deg t$, so dividing by $h\big(t(P)\big)$ yields

$$\frac{\deg f}{2\deg t} \ge \frac{h\big(f(P)\big)}{h\big(t(P)\big)} - \frac{c_2}{h\big(t(P)\big)} \qquad \text{for all } P \in C(K) \text{ with } f(P) \in R_S.$$

Finally, we make use of (B.3.5), which says that

$$\lim_{\substack{P\in C(K)\\ h(t(P))\to\infty}} \frac{h\big(f(P)\big)}{h\big(t(P)\big)} = \frac{\deg f}{\deg t}.$$

So if there were infinitely many points $P \in C(K)$ with $f(P) \in R_S$, then we could take a limit of such points with $h\big(t(P)\big) \to \infty$, in which case our inequality becomes

$$\frac{\deg f}{2\deg t} \ge \frac{\deg f}{\deg t}.$$

This contradiction shows that there are only finitely many points $P \in C(K)$ with $f(P) \in R_S$, which completes the proof of Siegel's theorem (D.9.1). □

Remark D.9.5. The proof of Roth's theorem being ineffective, the proof of Siegel's theorem that we have given is also ineffective. For many (but not all) curves, Baker's theorem, which provides lower bounds for linear forms of logarithms, can be used to make Siegel's theorem effective, as we now briefly describe. For further details, see Baker [1] or Serre [3].

Building on the method pioneered by Baker, Feldman [1] has shown that for all algebraic numbers α of degree $d \ge 3$, there exist two effectively computable constants $C = C(\alpha)$ and $\varepsilon = \varepsilon(\alpha) > 0$ such that for all rational numbers $p/q \in \mathbb{Q}$,

$$\left|\alpha - \frac{p}{q}\right| \ge \frac{C}{q^{d-\varepsilon}}.$$

Unfortunately, ε is quite small.

Baker's theorem says the following. Let $\alpha_1, \dots, \alpha_m$ be algebraic numbers. Then there is an effectively computable constant $c = c(\alpha_1, \dots, \alpha_m)$ such that for all integers $b_1, \dots, b_m$ with $\max_i |b_i| \le B$, either

$$\alpha_1^{b_1} \cdots \alpha_m^{b_m} = 1 \qquad \text{or} \qquad \left|\alpha_1^{b_1} \cdots \alpha_m^{b_m} - 1\right| \ge B^{-c}.$$

Note that the elementary Liouville inequality gives only

$$\left|\alpha_1^{b_1} \cdots \alpha_m^{b_m} - 1\right| \ge \exp(-c'B),$$

so Baker's theorem provides an exponential improvement.

Baker's result can be used to study the unit equation by choosing a set of generators $\alpha_1, \ldots, \alpha_m$ for the unit group and writing all units in the form $\alpha_1^{b_1} \cdots \alpha_m^{b_m}$. In this way Baker was able to prove effective bounds for the solutions to the unit equation (D.8.1), and hence for any Diophantine equation or problem that can be reduced to the solution of the unit equation. This includes, for example, the problem of integral solutions to the hyperelliptic equation $Y^2 = f(X)$ (cf. Theorem D.8.3), to so-called superelliptic equations of the form $Y^n = f(X)$ (see Exercise D.6), and to integral values of functions with at least three poles on curves of genus 0 (cf. Theorem D.8.4). The method also leads to an effective bound for the height of integral points on affine curves of genus 1, since these can be reduced to integral solutions of hyperelliptic equations. However, the method does not (at present) give an effective bound for integral points on an affine curve of genus 2 or greater, even if the curve is hyperelliptic, since integral points are not preserved by general birational transformations.

Exercises

D.1. (a) Prove that almost all real numbers (in the sense of measure theory) have approximation exponent 2. That is, prove that for every $\varepsilon > 0$, the set

$$\left\{ \alpha \in \mathbb{R} : \left| \frac{p}{q} - \alpha \right| \le \frac{1}{q^{2+\varepsilon}} \text{ has infinitely many solutions } \frac{p}{q} \in \mathbb{Q} \right\}$$

is a set of Lebesgue measure zero.

(b) More generally, let $F : \mathbb{N} \to \mathbb{R}$ be any real-valued function on the positive integers with the property that the series $\sum_{q \ge 1} q/F(q)$ converges. Prove that the set

$$\left\{ \alpha \in \mathbb{R} : \left| \frac{p}{q} - \alpha \right| \le \frac{1}{F(q)} \text{ has infinitely many solutions } \frac{p}{q} \in \mathbb{Q} \right\}$$

has measure zero.

Remark. Notice that Roth's theorem (D.2.1) says that the set in (a) contains no algebraic numbers. Lang has conjectured that the same is true for the set in (b). For example, this should be true for $F(q) = q^2(\log q)^{1+\varepsilon}$ for any $\varepsilon > 0$; but it is still an open question, even for $F(q) = q^2(\log q)^C$ with an arbitrarily large value of C.

D.2. (a) Prove that there are infinitely many rational numbers $p/q \in \mathbb{Q}$ satisfying

$$\left| \frac{p}{q} - \frac{1+\sqrt{5}}{2} \right| \le \frac{1}{\sqrt{5}q^2}.$$

(b) For any constant $\kappa > \sqrt{5}$, prove that there are only finitely many rational numbers $p/q \in \mathbb{Q}$ satisfying

$$\left|\frac{p}{q} - \frac{1+\sqrt{5}}{2}\right| \le \frac{1}{\kappa q^2}.$$

(c) If α is a real quadratic root of $aX^2+bX+c=0$, set $D := b^2-4ac$; prove that for all $\kappa > \sqrt{D}$, there are only finitely many rational numbers $p/q \in \mathbb{Q}$ satisfying

$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{\kappa q^2}.$$

D.3. Let $\alpha \in \mathbb{R}$ with $\alpha \notin \mathbb{Q}$.

(a) Prove that there are infinitely many rational numbers $p/q \in \mathbb{Q}$ satisfying

$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{\sqrt{5}q^2}.$$

(The previous exercise shows that $\sqrt{5}$ cannot in general be replaced by any larger constant.)

(b) Let $\beta = (1+\sqrt{5})/2$, and suppose that α cannot be written in the form $(a\beta+b)/(c\beta+d)$ with integers a, b, c, d satisfying $ad - bc = 1$. Prove that there are infinitely many rational numbers $p/q \in \mathbb{Q}$ satisfying

$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{\sqrt{8}q^2}.$$

D.4. Let $\alpha \in \bar{\mathbb{Q}}$ be an algebraic number of degree $d = [\mathbb{Q}(\alpha) : \mathbb{Q}] \ge 2$. Prove that there is a constant $c(\alpha) > 0$ such that

$$\left|\frac{p}{q} - \alpha\right| \ge \frac{c(\alpha)}{q^d} \quad \text{for all rational numbers } \frac{p}{q} \in \mathbb{Q}.$$

Find an explicit value for $c(\alpha)$ in terms of the height of α.

D.5. Use Liouville's theorem (D.1.2) or the previous exercise to prove that the number

$$\alpha = \sum_{n=0}^{\infty} \frac{1}{10^{n!}}$$

is transcendental over $\mathbb{Q}$. (Note that the exponent is n factorial.)

D.6. Let K be a number field and let $R_S \subset K$ be the ring of S-integers of K for some finite set of places S.

(a) Let $F(x,y) \in K[x,y]$ be a homogeneous polynomial of degree $d \geq 3$ with nonzero discriminant, and let $c \in K^*$. Without quoting Siegel's theorem (D.8.4 and D.9.1), use Roth's theorem directly to prove that the equation

$$F(x,y) = c$$

has only finitely many solutions $x, y \in R_S$. (Show that x/y closely approximates a root of $F(t,1)$. An alternative method is to reduce the problem to solving a certain collection of unit equations and apply Theorem D.8.1.) Equations of this form were first treated by Thue [1] in 1909.

(b) Let $f(x) \in K[x]$ be a polynomial of degree at least 2 with distinct roots (in $\bar{K}$), and let $n \geq 3$ be an integer. Mimic the proof of Theorem D.8.3 to prove that the equation $y^n = f(x)$ has only finitely many solutions $x, y \in R_S$.

D.7. (Leibniz's formula) Let $P, P' \in k[X_1, \ldots, X_m]$ be polynomials with coefficients in some ring k, and let $j_1, \ldots, j_m$ be an m-tuple of nonnegative integers. Prove that

$$\partial_{j_1 \ldots j_m}(PP') = \sum_{i_1+i_1'=j_1} \cdots \sum_{i_m+i_m'=j_m} (\partial_{i_1 \ldots i_m} P)(\partial_{i_1' \ldots i_m'} P').$$

D.8. Prove the estimate

$$e^t \leq 1 + t + t^2 \qquad \text{for all } |t| \leq 1$$

used in the proof of Lemma D.3.6.

D.9. Siegel's Lemma (D.4.1) says that a system of linear equations with integer coefficients has a solution of size at most $(N|A|)^{M/(N-M)}$. Prove that there is a solution whose size is at most $(\sqrt{N}|A|)^{M/(N-M)}$. Can you improve this bound further?

D.10. (a) Let $F \in \mathbb{Z}[X_1, \ldots, X_r]$ and $G \in \mathbb{Z}[Y_1, \ldots, Y_s]$ be polynomials that use different sets of variables. Prove that $|FG| = |F| \cdot |G|$, where recall that $|F|$ is the maximum absolute value of the coefficients of F.

(b) More generally, let K be a number field with ring of integers R_K, and let $F \in R_K[X_1, \ldots, X_r]$ and $G \in R_K[Y_1, \ldots, Y_s]$. Prove that $H_K(FG) = H_K(F)H_K(G)$.

(c) Give an example to show that (a) need not be true if we merely assume that F and G have coefficients in $\mathbb{Q}$.

D.11. Let $\phi_1, \ldots, \phi_k \in K(X)$ be rational functions over a field of characteristic 0. Suppose that there exists a nonzero generalized Wronskian of $\phi_1, \ldots, \phi_k$. Prove that $\phi_1, \ldots, \phi_k$ are linearly independent over K. (Note that this is precisely the part of Lemma D.6.1 that we left for you to do. Do not just quote Lemma D.6.1.)

D.12. Let $\alpha \in \bar{\mathbb{Q}}$, let $\varepsilon > 0$, and let $\mathcal{S}(\alpha, \varepsilon)$ denote the set of solutions $p/q \in \mathbb{Q}$ to the inequality
$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{q^{2+\varepsilon}}.$$
(a) If p_1/q_1 and p_2/q_2 are solutions in $\mathcal{S}(\alpha, \varepsilon)$ with $q_2 > q_1$, prove that
$$q_2 \ge \frac{1}{2} q_1^{1+\varepsilon}.$$
An inequality of this sort is called a *gap principle*, because it says that there are large gaps between solutions.
(b) Prove that for all $H_2 \ge H_1 > 2^{1/\varepsilon}$,
$$\#\left\{\frac{p}{q} \in \mathcal{S}(\alpha, \varepsilon) \,\middle|\, H_1 < q < H_2\right\} \le \log_{1+\varepsilon}\left(\frac{\log H_2}{\log 2^{-1/\varepsilon} H_1}\right).$$

D.13. Let $K/\mathbb{Q}$ be a number field; let $S \subset M_K$ be a finite set of absolute values on K, each extended in some way to $\bar{K}$; set $\operatorname{card}(S) = s$; let $\alpha \in \bar{K}$; and let $\varepsilon > 0$. Let $\xi : S \to [0, 1]$ be a function satisfying $\sum_{v \in S} \xi_v = 1$, and let $\mathcal{S}(K, S, \alpha, \varepsilon, \xi)$ be the set
$$\left\{\beta \in K \mid \|\beta - \alpha\|_v \le \frac{1}{H_K(\beta)^{(2+\varepsilon)\xi_v}} \quad \text{for all } v \in S\right\},$$
whose finiteness is proven in Theorem D.2.2.
(a) Prove that if $\beta_1, \beta_2 \in \mathcal{S}(K, S, \alpha, \varepsilon, \xi)$ satisfy $H_K(\beta_2) \ge H_K(\beta_1)$, then
$$H_K(\beta_2) \ge 2^{-([K:\mathbb{Q}]+s)} H_K(\beta_1)^{1+\varepsilon}.$$
This *gap principle* generalizes the previous exercise.
(b) Prove that for all constants $H_2 \ge H_1 > 2^{([K:\mathbb{Q}]+s)/\varepsilon}$,
$$\#\{\beta \in \mathcal{S}(K, S, \alpha, \varepsilon, \xi) : H_1 \le H_K(\beta) \le H_2\} \le \log_{1+\varepsilon}\left(\frac{\log H_2}{\log 2^{-([K:\mathbb{Q}]+s)/\varepsilon} H_1}\right).$$

D.14. Let $K/\mathbb{Q}$ be a number field; let $S \subset M_K$ be a finite set of absolute values on K, each extended in some way to $\bar{K}$; let $\alpha \in \bar{K}$; and let $\varepsilon > 0$. Let $\mathcal{S}(K, S, \alpha, \varepsilon)$ be the set
$$\mathcal{S}(K, S, \alpha, \varepsilon) = \left\{\beta \in K \mid \prod_{v \in S} \min\{\|\beta - \alpha\|_v, 1\} \le \frac{1}{H_K(\beta)^{2+\varepsilon}}\right\}$$
considered in Theorem D.2.1, and for any function
$$\xi : S \to [0, 1] \quad \text{satisfying} \quad \sum_{v \in S} \xi_v = 1,$$
let $\mathcal{S}(K, S, \alpha, \varepsilon, \xi)$ be the set
$$\mathcal{S}(K, S, \alpha, \varepsilon, \xi) = \left\{\beta \in K \mid \|\beta - \alpha\|_v \le \frac{1}{H_K(\beta)^{(2+\varepsilon)\xi_v}} \quad \text{for all } v \in S\right\}$$
considered in Theorem D.2.2. Prove that
$$\#\mathcal{S}(K, S, \alpha, \varepsilon) \le 4^{\#S} \sup_{\xi} \#\mathcal{S}\left(K, S, \alpha, \frac{1}{2}\varepsilon, \xi\right),$$
where the supremum is over all $\xi : S \to [0, 1]$ with $\sum_{v \in S} \xi_v = 1$. (Notice that this provides a quantitative proof of the reduction lemma (D.2.2.1), which says that Theorem D.2.2 implies Theorem D.2.1.)

D.15. Let $K/\mathbb{Q}$ be a number field; let $S \subset M_K$ be a finite set of absolute values on K, each extended in some way to $\bar{K}$; let $\alpha \in \bar{K}$; and let $\varepsilon > 0$. Let $\mathcal{S}(K,S,\alpha,\varepsilon)$ be the set

$$\mathcal{S}(K,S,\alpha,\varepsilon) = \left\{ \beta \in K \mid \prod_{v \in S} \min\{\|\beta-\alpha\|_v, 1\} \le \frac{1}{H_K(\beta)^{2+\varepsilon}} \right\}$$

considered in Theorem D.2.1.

(a) Prove that there are constants C_1 and C_2, depending only on $[K(\alpha):\mathbb{Q}]$ and ε, such that

$$\#\{\beta \in \mathcal{S}(K,S,\alpha,\varepsilon) \mid H_K(\beta) \ge \max(H_K(\alpha),2)^{C_1}\} \le 4^{\#S} C_2.$$

(b) Find explicit expressions for C_1 and C_2 in terms of $[K(\alpha):\mathbb{Q}]$ and ε.

D.16. Let $C/\mathbb{Q}$ be a smooth projective curve of genus 1, and suppose that $C(\mathbb{Q})$ is an infinite set. Fix a nonconstant function $f \in \mathbb{Q}(C)$, and for each point $P \in C(\mathbb{Q})$, write

$$f(P) = \frac{a_P}{b_P} \in \mathbb{Q}$$

as a fraction in lowest terms. Prove that

$$\lim_{P \in C(\mathbb{Q}),\, h(f(P)) \to \infty} \frac{\log |a_P|}{\log |b_P|} = 1.$$

Notice that this says that the numerator and the denominator of $f(P)$ have approximately the same number of digits. It greatly strengthens Siegel's theorem (D.9.1), which in this situation would merely say that there are finitely many $P \in C(\mathbb{Q})$ with $b_P = 1$ (i.e., with $f(P) \in \mathbb{Z}$).

D.17. Generalize the previous exercise as follows. Let C/K be a smooth projective curve of genus 1 such that $C(K)$ is an infinite set, and fix a nonconstant function $f \in K(C)$. Also, let h_C be a height function on C with respect to some fixed ample divisor, and let S be a finite set of places of K. Prove that

$$\lim_{P \in C(K),\, h_C(P) \to \infty} \frac{\sum_{v \in S} \left|\log \|f(P)\|_v\right|}{h_C(P)} = 0.$$

D.18. *Continued fractions.* To ease notation, for $a_0 \in \mathbb{R}$ and $a_1, \ldots, a_n > 0$ we set

$$[a_0, a_1, \ldots, a_n] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{a_n}}}$$

To every real number x, we associate a sequence of integers a_n (with $a_1, \ldots, a_n > 0$) and an auxiliary real sequence x_n defined by

$$a_0 = [x],\ x_0 = x \qquad \text{and} \qquad x_{n+1} = \frac{1}{x_n - a_n},\ a_{n+1} = [x_{n+1}],$$

with the convention that the sequence terminates if x_n is an integer (which occurs only when $x \in \mathbb{Q}$). Further, define the nth *convergent* of x to be the rational number

$$\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n].$$

(a) Prove that $x = [a_0, \ldots, a_{n-1}, x_n]$.

(b) Verify the following useful recursion formulas:

$$\begin{aligned} p_{n+1} &= a_{n+1}p_n + p_{n-1}, \\ q_{n+1} &= a_{n+1}q_n + q_{n-1}, \\ q_n p_{n-1} - p_n q_{n-1} &= (-1)^n, \\ q_n p_{n-2} - p_n q_{n-2} &= (-1)^{n-1} a_n. \end{aligned}$$

(c) Let $x \notin \mathbb{Q}$ be an irrational number. Show that the convergents p_n/q_n provide "good" approximations to x in the sense that

$$\frac{1}{q_n(q_n + q_{n+1})} < \left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}.$$

Prove that they provide the "best" approximations to x in the sense that

$$q_n \left| x - \frac{p_n}{q_n} \right| < \min_{\substack{p/q \in \mathbb{Q} \\ q < q_n}} \left\{ q \left| x - \frac{p}{q} \right| \right\}.$$

(d) (Lagrange) Show that the sequence a_n eventually becomes periodic if and only if x is quadratic, i.e., x is the root of an irreducible quadratic polynomial $x^2 + Ax + B \in \mathbb{Q}[x]$.

(e) Show that the following are equivalent:

(i) The sequence a_n is bounded.

(ii) There exists a constant $C = C(x)$ such that $\left| x - \frac{p}{q} \right| \geq C/q^2$ for all rational numbers p/q.

It is suspected that a_n is unbounded for all algebraic numbers of degree at least 3, but this is not known for be true for even a single such number!

PART E

Rational Points on Curves of Genus at Least 2

L'arithmétique
Est une mécanique
Qui donne la colique
Aux catholiques,
Le mal au cœur
Aux enfants de chœurs,
Et le mal au nez
Aux curés.
Comptine du Bourbonnais

Let $K/\mathbb{Q}$ be a number field and C/K a curve of genus g defined over K. If $g = 0$, then we have seen in Part A.4.3 that $C \cong \mathbb{P}^1$ (over $\bar{K}$), so the set of K-rational points $C(K)$ on C is either empty or equal to $\mathbb{P}^1(K)$. If $g = 1$, then C is an elliptic curve and has the structure of an abelian variety of dimension 1. The Mordell–Weil theorem (C.0.1), proven for $K = \mathbb{Q}$ by Mordell [1] in 1922 and in general by Weil [1] in 1928, says that if $C(K)$ is nonempty, then it is a finitely generated abelian group. In particular, if $g \leq 1$, then it frequently happens that $C(K)$ is an infinite set, and this will always be true for $C(L)$ for some finite extension L/K. In stark contrast stands the following result, conjectured by Mordell [1] in his 1922 paper and first proven by Faltings [1] in 1983.

Theorem E.0.1. (Faltings [1]) *Let K be a number field, and let C/K be a curve of genus $g \geq 2$. Then $C(K)$ is finite.*

Faltings' [1] proof of Theorem E.0.1 in 1983 used a variety of advanced techniques from modern algebraic geometry, including tools such as moduli schemes and stacks, semistable abelian schemes, and p-divisible groups. Vojta [2] then came up with an entirely new proof of Faltings' theorem using ideas whose origins lie in the classical theory of Diophantine approximation. However, in order to obtain the precise estimates needed for the delicate arguments involved, he made use of Arakelov arithmetic intersection theory and the deep and technical Riemann–Roch theorem for arithmetic threefolds proven by Gillet and Soulé. Faltings [2] then simplified Vojta's proof by eliminating the use of the Gillet–Soulé theorem and proving a "product

lemma" especially well suited to induction. This allowed Faltings to generalize Vojta's result to prove a conjecture of Lang concerning rational and integral points on subvarieties of abelian varieties (see Part F for further comments). However, Faltings' proof, which uses arithmetic intersection theory and heights defined via differential geometric considerations, is far from elementary.

Bombieri [1] has combined Faltings' generalization with Vojta's original proof and with other simplifications of his own to give a comparatively elementary proof of the original Mordell conjecture. In addition to the Mordell–Weil theorem (C.0.1) for the Jacobian of C, the tools used in Bombieri's proof fall broadly into the following three areas:

(i) *Geometric Tools*
The Riemann–Roch theorem for surfaces, or more precisely, for the product $C \times C$ of a curve with itself. The theory of curves, Jacobians, and theta divisors. Aside from the necessity of keeping track of fields of definition, this material dates from the nineteenth century. Especially useful is Riemann's theorem describing the intersection of a curve with a translation of the theta divisor. This material is surveyed in Part A.

(ii) *Height Functions*
Weil height functions associated to divisor classes. Canonical height functions on abelian varieties and their associated quadratic forms. The theory of height functions is, in essence, a tool for translating geometric information, in the form of relations between divisor classes, into arithmetic information about points. The theory of what are now known as Weil heights was developed during the 1940s and 50s (see, e.g., Weil [4] or Northcott [1, 2]), and canonical heights were constructed by Néron [2] and Tate [unpublished] in the mid-1960s. We should also mention Mumford's application [1] of canonical heights to Mordell's conjecture in 1965, since several of Mumford's ideas play a crucial role in Vojta's proof. We have covered this material in Part B.

(iii) *Diophantine Approximation*
The classical theory of Diophantine approximation asks how closely an irrational quantity can be approximated by a rational quantity. Proofs in this subject follow a basic pattern: (1) Construction of an auxiliary function using Siegel's lemma. (2) An elementary upper bound, essentially obtained from the triangle inequality. (3) A nonvanishing result, such as Dyson's lemma or Roth's lemma. (4) A lower bound, obtained, via the product formula, from the fact that 1 is the smallest positive integer. We have discussed the theory of Diophantine approximation in Part D.

All four of the Diophantine approximation steps appear in the proof of Mordell's conjecture, and anyone familiar with the proof of Roth's theorem in Part D will have no trouble picking out each step as we go along. However, one new feature to observe is that rather

than studying approximations to a point, we instead look at points that approximate a certain carefully chosen curve representing a particular divisor class in $C \times C$. We will use the Riemann–Roch theorem on $C \times C$ to tell us that the particular divisor class actually contains a curve (i.e., the divisor class is effective).

In some sense, all of the tools needed for the proof of Mordell's conjecture were available by 1965, when Mumford's paper [1] appeared. In the remainder of this chapter we will present the proof of Mordell's conjecture as given in Bombieri's paper [1]. The following material will be used in the proof, and may thus be considered prerequisite for reading this part.

Part A, Sections 1–8.
Part B, Sections 1–5 and 7.
Part C, Sections 1–2.
Part D, Sections 1–7.

E.1. Vojta's Geometric Inequality and Faltings' Theorem

In this section we will describe an inequality due to Vojta and show how it leads, via an elementary geometric argument, to a proof of Faltings' theorem (E.0.1). Most of the remainder of this chapter will then be devoted to proving Vojta's inequality.

We begin by setting a little notation, which will remain fixed throughout.

- K — a number field.
- C/K — a smooth projective curve of genus $g \geq 2$ defined over K. We will assume that $C(K)$ is nonempty, since otherwise Theorem E.0.1 is trivially true.
- J/K — the Jacobian variety of C.
- Θ — the theta divisor on J. Recall (see Corollary A.8.2.3) that Θ is an ample divisor on J.
- $|\cdot|$ — the norm on $J(\bar{K})$ associated to the canonical height relative to Θ. In other words, $|x|^2 = \hat{h}_{J,\Theta}(x)$. We recall (Proposition B.5.3) that $|\cdot|$ extends to a positive definite quadratic form on the vector space $J(K) \otimes \mathbb{R}$.
- $\langle \cdot, \cdot \rangle$ — the bilinear form on $J(\bar{K})$ associated to the canonical height relative to the divisor Θ. In other words,
  $$\langle x, y \rangle = \frac{1}{2}\left(|x+y|^2 - |x|^2 - |y|^2\right).$$
  The inner product $\langle \cdot, \cdot \rangle$ extends to a Euclidean inner product on the vector space $J(K) \otimes \mathbb{R}$.

We choose a rational point in $C(K)$ and use it to fix an embedding (defined over K)

$$C \hookrightarrow J.$$

Having done this, we can talk about the norm $|z|$ and inner product $\langle z, w\rangle$ for points $z, w \in C(\bar{K})$. We are now ready for Vojta's inequality, whose innocuous statement belies its far-reaching consequence.

Theorem E.1.1. Vojta's Inequality. (Vojta [1], Bombieri [1]) *With notation as above, there are constants $\kappa_1 = \kappa(C)$ and $\kappa_2 = \kappa_2(g)$ such that if $z, w \in C(\bar{K})$ are two points satisfying*

$$|z| \geq \kappa_1 \quad \text{and} \quad |w| \geq \kappa_2 |z|, \quad \text{then} \quad \langle z, w\rangle \leq \frac{3}{4}|z|\,|w|.$$

Before using Vojta's inequality to prove Faltings' theorem in full generality, let us look at a special (but still highly nontrivial) case. Suppose that $J(K)$ is generated by a single point x_0 of infinite order, so $J(K) = \mathbb{Z}x_0$. Then $C(K)$ consists of those multiples nx_0 that happen to lie on C. To prove that $C(K)$ is finite, we must show that there are only finitely many n's with $nx_0 \in C$.

We give a proof by contradiction, so we suppose to the contrary that infinitely many multiples of x_0 lie on C. Replacing x_0 by $-x_0$ if necessary, we may assume that infinitely many positive multiples of x_0 lie on C. In particular, we can find some multiple $n_1 x_0 \in C$ satisfying $n_1 > \kappa_1/|x_0|$. (Here κ_1 is the constant in Vojta's inequality. Note that $|x_0| > 0$, since x_0 is a nontorsion point.) Similarly, we can find a multiple $n_2 x_0 \in C$ satisfying $n_2 > \kappa_2 n_1$. This means that

$$|n_1 x_0| = n_1|x_0| > \kappa_1 \quad \text{and} \quad |n_2 x_0| = \frac{n_2}{n_1}|n_1 x_0| > \kappa_2 |n_1 x_0|,$$

so we can apply Vojta's inequality with $z = n_1 x_0$ and $w = n_2 x_0$ to conclude that

$$\langle n_1 x_0, n_2 x_0\rangle \leq \frac{3}{4}|n_1 x_0|\,|n_2 x_0|.$$

$$n_1 n_2 \langle x_0, x_0\rangle \leq \frac{3}{4} n_1 n_2 |x_0|^2.$$

But $\langle x_0, x_0\rangle = |x_0|^2 > 0$ and $n_1, n_2 \geq 1$, so this is a contradiction, which completes the proof in this special case that Vojta's inequality (E.1.1) implies Faltings' theorem (E.0.1).

The proof of the general case is similar; we need merely deal with the fact that $J(K)$ may have rank greater than 1. The key is to understand geometrically what Vojta's inequality says about the image of the set $C(K)$ in the Euclidean vector space $J(K) \otimes \mathbb{R}$.

Proposition E.1.2. *Let C/K be a curve of genus at least 2 defined over a number field K. Then Vojta's inequality (E.1.1) implies Faltings' theorem (E.0.1) that $C(K)$ is finite.*

PROOF. First we observe that the kernel of the map $J(K) \to J(K) \otimes \mathbb{R}$ is the torsion subgroup $J(K)_{\text{tors}}$, which is finite (Theorem C.0.1). So in order

to prove that $C(K)$ is finite, it suffices to show that the image of $C(K)$ in $J(K) \otimes \mathbb{R}$ is finite. By abuse of notation, we will identify $C(K)$ with its image.

The bilinear form $\langle \cdot, \cdot \rangle$ makes $J(K) \otimes \mathbb{R}$ into a finite-dimensional Euclidean space, so for any two points $x, y \in J(K) \otimes \mathbb{R}$ we can define the "angle" $\theta(x,y)$ between x and y in the usual way,

$$\cos\theta(x,y) = \frac{\langle x,y \rangle}{|x|\,|y|}, \qquad 0 \le \theta(x,y) \le \pi.$$

For any point x_0 and any angle θ_0, we consider the cone with interior angle $2\theta_0$ whose central axis is the ray from 0 through x_0,

$$\Gamma_{x_0,\theta_0} = \bigl\{ x \in J(K) \otimes \mathbb{R} \bigm| \theta(x, x_0) < \theta_0 \bigr\}.$$

For example, if $J(K)$ has rank 2, then one of these cones would look like the sector illustrated in Figure E.1.

A Cone Γ_{x_0,θ_0} in $J(K) \otimes \mathbb{R}$
Figure E.1

We are going to use Vojta's inequality to show that if θ_0 is small enough, then every cone Γ_{x_0,θ_0} intersects $C(K)$ in only finitely many points. (For example, we will show this is true if $\theta_0 = \pi/12$.) To see this, suppose that we have a cone satisfying

$$\#\bigl(\Gamma_{x_0,\theta_0} \cap C(K)\bigr) = \infty.$$

Since $J(K)$, and a fortiori $C(K)$, contains only finitely many points of bounded norm, we can find a $z \in \Gamma_{x_0,\theta_0} \cap C(K)$ with $|z| \geq \kappa_1$, and then we can find a $w \in \Gamma_{x_0,\theta_0} \cap C(K)$ with $|w| \geq \kappa_2 |z|$. Here κ_1 and κ_2 are the constants appearing in Vojta's inequality (E.1.1). Then Vojta's inequality tells us that

$$\langle z, w \rangle \leq \frac{3}{4} |z|\,|w|,$$

or equivalently,

$$\cos\theta(z,w) \leq \frac{3}{4}.$$

This estimate captures the essence of Vojta's inequality; it says that the angle between the points z and w cannot be too small. For example, it implies that

$$\theta(z,w) \geq \cos^{-1}\frac{3}{4} > \frac{\pi}{6}.$$

But by assumption, both z and w are in the cone Γ_{x_0,θ_0}, so the angle between them is less than $2\theta_0$. We have thus shown that

$$\#\Gamma_{x_0,\theta_0} = \infty \qquad \text{implies} \qquad 2\theta_0 > \theta(z,w) \geq \frac{\pi}{6}.$$

This is equivalent to the statement that

$$\Gamma_{x_0,\pi/12} \cap C(K) \text{ is finite for every } x_0 \in J(K) \otimes \mathbb{R}.$$

In order to complete the proof that $C(K)$ is finite, we now need merely observe that it is possible to cover $J(K) \otimes \mathbb{R}$ by finitely many cones of the form $\Gamma_{x_0,\pi/12}$. If this is not immediately obvious, consider the intersection of such cones with the unit sphere

$$S = \bigl\{ x \in J(K) \otimes \mathbb{R} \bigm| |x| = 1 \bigr\} \subset J(K) \otimes \mathbb{R}.$$

Clearly,

$$S = \bigcup_{x \in S} \bigl(\Gamma_{x,\pi/12} \cap S \bigr),$$

since $x \in \Gamma_{x,\pi/12}$. Further, each set $\Gamma_{x,\pi/12} \cap S$ is an open subset of S, and we know that S is compact, from which it follows that S is covered by finitely many of the $\Gamma_{x,\pi/12} \cap S$'s (see the comments at the end of this chapter and Exercise E.1 for a more effective argument). Since the Γ's are cones, we conclude that the same finite set of Γ's will cover $J(K) \otimes \mathbb{R}$. This completes the proof that Vojta's inequality implies that $C(K)$ is finite. □

We now take up the formidable task of proving Vojta's inequality (E.1.1).

E.2. Pinning Down Some Height Functions

One of the failings of the theory of Weil heights is that they are defined only up to $O(1)$, and the $O(1)$'s depend on the particular embeddings chosen for each very ample divisor. Thus if $D \in \mathrm{Div}(V)$ is a divisor and we want to choose a particular height function $h_{V,D}$, we need to make the following choices:

[1] Choose very ample divisors D_1 and D_2 with $D = D_1 - D_2$. Generally one does this by choosing a divisor E such that E and $D+E$ are both very ample, and then taking $D_1 = D + E$ and $D_2 = E$.
[2] Choose embeddings $\phi_1 : V \to \mathbb{P}^n$ and $\phi_2 : V \to \mathbb{P}^m$ corresponding respectively to D_1 and D_2.
[3] Set $h_{V,D}(P) = h\big(\phi_1(P)\big) - h\big(\phi_2(P)\big)$.

We could, a priori, decide to make these choices for every divisor on every variety, thereby fixing a particular height function $h_{V,D}$ for each divisor D on each variety V. Then the Height Machine works, but as already noted, it works only up to bounded functions. For example, if we fix our heights and if $\psi : W \to V$ is a morphism, then the function

$$\big(h_{W,\psi^*D} - h_{V,D} \circ \psi\big) : W(\bar{K}) \longrightarrow \mathbb{R}$$

is bounded, but the bound depends not only on ψ, but also on the particular height functions we have chosen for D and ψ^*D. In proving Vojta's inequality (E.1.1) we will need to choose our height functions in a uniform manner so as to be able to determine how the $O(1)$'s depend on certain parameters.

Let us be a bit more precise. We will be using height functions on $C \times C$ corresponding to *Vojta divisors* $\Omega = \Omega(d_1, d_2, d)$ depending on three integer parameters d_1, d_2, d. What we are going to do is write the height function $h_{C\times C,\Omega}$ as a linear combination

$$h_{C\times C,\Omega(d_1,d_2,d)} = c_1 d_1 h_1 + c_2 d_2 h_2 + c_3 d h_3,$$

where c_1, c_2, c_3 are fixed constants and h_1, h_2, h_3 are fixed height functions corresponding to particular maps of $C \times C$ into projective space. The crucial property to keep in mind is that the c_i's and h_i's do not depend on d_1, d_2, d, so as we vary d_1, d_2, d, we have complete control over the variation of the height $h_{C\times C,\Omega(d_1,d_2,d)}$. The reason we need this property is that at the final step in the proof of Vojta's inequality, we will choose d_1, d_2, and d to depend on the heights of the two points $z, w \in C(\bar{K})$, so it will be vital that all of the $O(1)$ and c_i constants floating around be independent of d_1, d_2, d.

With this motivation, we are now going to set some notation and fix some embeddings and height functions. We start by fixing a divisor $A \in$

$\mathrm{Div}(C)$ of degree 1. In order to link up with (A.8.2.1), we will take A to satisfy the condition

$$(2g-2)A \sim \mathcal{K}_C, \tag{5}$$

where $\mathcal{K}_C$ is the canonical divisor class on C. It is always possible to find such an A, since the group $J(\bar{K})$ is divisible. Of course, A need not be defined over the base field K, but it will be defined over a finite extension of K, and for the proof of Vojta's inequality we are always free to replace K by a finite extension.

We use the divisor A to define an embedding

$$j_A : C \longrightarrow J, \qquad x \longmapsto \mathrm{Cl}\big((x) - A\big).$$

We further let

$$\Theta_A = \underbrace{j_A(C) + j_A(C) + \cdots + j_A(C)}_{g-1 \text{ copies}}$$

be a theta divisor on J. (This is actually a translation of the theta divisors considered in Section A.8.) The fact that we chose A to satisfy (5) implies the following useful relations, which will simplify our later calculations.

Lemma E.2.1. *With notation as above, we have:*
(a) *Θ_A is a symmetric divisor, that is, $\Theta_A^- = \Theta_A$.*
(b) *$j_A^*\Theta_A \sim gA$.*
(c) *Let $s_{12}, p_1, p_2 : J \times J \to J$ be the summation and projection maps, respectively, and let $\Delta \in \mathrm{Div}(C \times C)$ be the diagonal divisor. Then*

$$(j_A \times j_A)^*(s_{12}^*\Theta_A - p_1^*\Theta_A - p_2^*\Theta_A) \sim -\Delta + p_1^*A + p_2^*A.$$

PROOF. All of these relations for Θ_A follow easily from our choice of A satisfying (5), since the corresponding relations for Θ were proven in Theorem A.8.2.1. Thus in the notation of Theorem A.8.2.1, we fix a point $P_0 \in C$ and use it to define an embedding $j : C \to J$ via $j(P) = \mathrm{Cl}((P) - (P_0))$. Notice that $j_A(x) = j(x) - j(A)$, and hence

$$\begin{aligned}\Theta_A &= j_A(C) + \cdots + j_A(C) \\ &= j(C) + \cdots + j(C) - (g-1)j(A) = \Theta - j\big((g-1)A\big).\end{aligned}$$

Using this, (5), and (A.8.2.1(i)), we complete the proof of (a) by computing

$$\begin{aligned}(\Theta_A)^- &= \Theta^- + j\big((g-1)A\big) \sim \Theta - j(\mathcal{K}_C) + j\big((g-1)A\big) \\ &\sim \Theta - j\big((g-1)A\big) = \Theta_A.\end{aligned}$$

Similarly, using (5) and (A.8.2.1(ii)), we find that

$$\begin{aligned} j_A^*\Theta_A &= j_A^*\left(\Theta - j((g-1)A)\right) \\ &= j^*\left(\Theta - j((g-2)A)\right) \\ &\sim g(P_0) - j((g-2)A) + j(\mathcal{K}_C) \\ &\sim g(P_0) - (g-2)(A-(P_0)) + (\mathcal{K}_C - (2g-2)(P_0)) \\ &= -(g-2)A + \mathcal{K}_C \\ &\sim gA. \end{aligned}$$

This gives (b). Finally, (c) follows from a similar calculation using (5) and (A.8.2.1(iii)), a task that we leave to the reader. □

In order to ease notation, we will often drop the j_A and just treat C as a subvariety of J. Further, the canonical height $\hat{h}_J$ on J will always refer to the canonical height with respect to the divisor Θ_A, and similarly for the corresponding norm $|\cdot|^2 = \hat{h}_{J,\Theta}$ and inner product $\langle\cdot,\cdot\rangle$. For example, when we refer to the canonical height $\|x\|^2$ of a point $x \in C(\bar{K})$, we really mean the canonical height $\hat{h}_{J,\Theta_A}$ of the point $j_A(x)$.

We also use A to give various embeddings into projective space. We choose an integer N such that the divisor NA is very ample, and fix an embedding

$$\phi_{NA} : C \longrightarrow \mathbb{P}^n$$

corresponding to NA. (For example, Corollary A.4.2.4 implies that $N = 2g+1$ is large enough.)

Remark.E.2.2. Since $x_0, \ldots, x_n$ form a basis of the space of sections of $\mathcal{O}(NA)$, it follows that $\phi_{NA}(C)$ is not contained in any hyperplane of $\mathbb{P}^n$. It will be convenient to choose the projective coordinates $x_0, \ldots, x_n$ such that the following properties are satisfied:

(i) The intersection of $\phi_{NA}(C)$ with the codimension-2 subspaces $x_i = x_j = 0$ is empty, and hence the projections $(x_0, \ldots, x_n) \mapsto (x_i, x_j)$ from $\phi_{NA}(C)$ to $\mathbb{P}^1$ all have degree N.

(ii) The projections $(x_0, \ldots, x_n) \mapsto (x_i, x_j, x_\ell)$ from $\phi_{NA}(C)$ to $\mathbb{P}^2$ are birational, so that $k(C) = k(x_j/x_i, x_\ell/x_i)$ and x_ℓ/x_i is integral of precise degree N over x_j/x_i.

In fact, "most" linear combinations of $x_0, \ldots, x_n$ will satisfy these properties (see Exercise E.10).

Next we look at divisors on $C \times C$. For example, we have the "slices" $A \times C$ and $C \times A$, and the diagonal Δ. Since A is ample on C, it is clear that the divisor $(A \times C) + (C \times A)$ is ample on $C \times C$, so if we choose a sufficiently large integer M, then the divisor

$$B = (M+1)(A \times C) + (M+1)(C \times A) - \Delta \in \operatorname{Div}(C \times C)$$

will be very ample by (A.3.2.3) (see Exercise E.12). Having chosen s, we fix an embedding

$$\phi_B : C \times C \longrightarrow \mathbb{P}^m$$

corresponding to a basis for the linear system $L(B)$. We will use homogeneous coordinates $[y_0, \dots, y_m]$ on $\mathbb{P}^m$, and by abuse of notation, we will also write $y_0, \dots, y_m$ for the corresponding sections of $\mathcal{O}(B)$. [More properly, y_i is a global section of $\mathcal{O}_{\mathbb{P}^m}(1)$, and we should write $\phi_B^* y_i$ for the corresponding section of $\mathcal{O}(B)$.] Now we can fix a particular height function for the divisor B by the formula

$$h_{C\times C,B}(z,w) = h\bigl(\phi_B(z,w)\bigr),$$

or equivalently in terms of coordinate functions,

$$h_{C\times C,B} = h\bigl([y_0, \dots, y_m]\bigr).$$

Similarly, for any integer $d \geq 1$ we can use linearity to fix a height for the divisor dB by setting

$$h_{C\times C,dB} = dh_{C\times C,B}.$$

One other comment, of a geometric nature, is needed. It is clear that any monomial in $y_0, \dots, y_m$ of degree d will be a global section of $\mathcal{O}(dB)$. What is also true is that if d is chosen sufficiently large (how large depends on C and B), then the monomials of degree d in $y_0, \dots, y_m$ generate the space of global sections to $\mathcal{O}(dB)$. In fancy language, this is a consequence of the fact that for sufficiently large d, the d-uple embedding of a smooth (or even just normal) variety is projectively normal. For a proof, see Theorem A.3.2.5 or Mumford [4, Chapter 6, (6.10) Theorem, page 102], or Hartshorne [1, Exercise II.5.14]. The same arguments apply to a product of projective varieties, for example to $C \times C \hookrightarrow \mathbb{P}^n \times \mathbb{P}^n$. When δ_1 and δ_2 are large enough the space of sections to $\mathcal{O}(\delta_1(NA \times C) + \delta_2(C \times NA))$ is generated by monomials of bidegree (δ_1, δ_2) in $[x_0, \dots, x_n; x'_0, \dots, x'_n]$.

We next use two copies of ϕ_{NA} to create a product embedding,

$$\phi_{NA} \times \phi_{NA} : C \times C \longrightarrow \mathbb{P}^n \times \mathbb{P}^n.$$

We will use bihomogeneous coordinates $[x_0, \dots, x_n; x'_0, \dots, x'_n]$ on $\mathbb{P}^n \times \mathbb{P}^n$. Let δ_1 and δ_2 be (large) integers. If we compose $\phi_{NA} \times \phi_{NA}$ with the δ_1-uple embedding of the first $\mathbb{P}^n$ and the δ_2-uple embedding of the second $\mathbb{P}^n$, and then compose with the Segre embedding (Example A.1.2.6(b)), the composition

$$C \times C \xrightarrow{\phi_{NA}\times\phi_{NA}} \mathbb{P}^n \times \mathbb{P}^n \xrightarrow{(d_1\text{-uple})\times(d_2\text{-uple})} \mathbb{P} \times \mathbb{P} \xrightarrow{\text{Segre}} \mathbb{P}$$

is associated to the divisor

$$\delta_1(NA \times C) + \delta_2(C \times NA).$$

So for all positive integers δ_1 and δ_2, we can fix a height function for the divisor $\delta_1(NA \times C) + \delta_2(C \times NA)$ by setting

$$\begin{aligned} h_{C\times C,\delta_1(NA\times C)+\delta_2(C\times NA)}(z,w) &= \delta_1 h_{C,NA}(z) + \delta_2 h_{C,NA}(w) \\ &= \delta_1 h\big(\phi_{NA}(z)\big) + \delta_1 h\big(\phi_{NA}(w)\big). \end{aligned}$$

In terms of coordinate functions, this becomes

$$h_{C\times C,\delta_1(NA\times C)+\delta_2(C\times NA)} = \delta_1 h\big([x_0,\dots,x_n]\big) + \delta_2 h\big([x'_0,\dots,x'_n]\big).$$

We also want to observe that if δ_1 and δ_2 are sufficiently large, then every global section of $\mathcal{O}\big(\delta_1(NA\times C)+\delta_2(C\times NA)\big)$ is the pullback to $C\times C$ via ψ of a bihomogeneous polynomial of bidegree (δ_1,δ_2) in the variables $x_0,\dots,x_n,x'_0,\dots,x'_n$. As above, this is a consequence of the fact that a large d-uple embedding of a normal variety is projectively normal.

With these preliminaries out of the way, we are ready to define the divisors and the height functions that will occupy most of our attention. For any given integers d_1, d_2, and d, we define

$$\Omega = \Omega(d_1,d_2,d) = (d_1-d)(A\times C) + (d_2-d)(C\times A) + d\Delta \in \operatorname{Div}(C\times C),$$

where $A \in \operatorname{Div}(C)$ is the divisor of degree 1 fixed above and Δ is the diagonal divisor. We say that Ω is a *Vojta divisor* if d_1, d_2, and d are positive integers satisfying the inequalities

$$gd^2 < d_1 d_2 < g^2 d^2. \tag{6}$$

(The first inequality will guarantee that $\Omega(d_1,d_2,d)$ will be linearly equivalent to an effective divisor, whereas the second inequality will ensure that the associated Néron–Tate quadratic form will not be positive definite.) We will generally assume that (6) holds and that d_1, d_2, and d are all divisible by the integer N that we fixed earlier. We then set

$$\delta_1 = \frac{d_1 + Md}{N} \qquad \text{and} \qquad \delta_2 = \frac{d_2 + Md}{N},$$

where M is as specified above. Finally, we will assume that d_1, d_2, and d are chosen sufficiently large so that the global sections of $\mathcal{O}(dB)$ are generated by monomials of degree d in $y_0,\dots,y_m$, and so that the global sections of $\mathcal{O}\big(\delta_1(NA\times C)+\delta_2(C\times NA)\big)$ are generated by monomials of bidegree (δ_1,δ_2) in $x_0,\dots,x_n,x'_0,\dots,x'_n$.

The integers that we eventually choose will satisfy $d_1 > d > d_2$, so the Vojta divisor itself is not positive, although we will see using Riemann–Roch that it is linearly equivalent to a positive divisor. In any case, we want to choose a height function on $C\times C$ corresponding to $\Omega(d_1,d_2,d)$ in such a way that we can explicitly keep track of the dependence on d_1, d_2,

and d. Using the height functions described above, this is easy. We just write Ω as the difference

$$\Omega(d_1, d_2, d) = \{\delta_1(NA \times C) + \delta_2(C \times NA)\} - dB$$

of very ample divisors. Then from above we see that we may take

$$h_{C\times C,\Omega(d_1,d_2,d)}(z,w) = \delta_1 h_{C,NA}(z) + \delta_2 h_{C,NA}(w) - d h_{C\times C,B}(z,w). \quad (7)$$

Alternatively, in terms of coordinates, this is just

$$h_{C\times C,\Omega(d_1,d_2,d)} = \delta_1 h\big([x_0,\dots,x_n]\big) + \delta_2 h\big([x'_0,\dots,x'_n]\big) - d h\big([y_0,\dots,y_m]\big). \quad (8)$$

From now on, whenever we talk about the height on $C \times C$ relative to a Vojta divisor $\Omega(d_1, d_2, d)$, we will mean the particular height function specified by either of the equivalent formulas (7) or (8).

We conclude this section with a short table of constants that will appear frequently during the proof of Vojta's inequality. The reader is well advised to refer to this table until he/she is thoroughly acquainted with this notation.

γ, ε, ν	small positive constants
$c_1, c_2, \dots$	constants that depend on C, A, and the choice of various height functions
M	an integer chosen large enough so that B is a very ample divisor on $C \times C$. This integer is fixed once and for all at the start of the proof
N	an integer chosen large enough so that NA is a very ample divisor on C. This integer is fixed once and for all at the start of the proof.
d_1, d_2, d	large integers, divisible by N, assumed to satisfy $gd^2 < d_1 d_2 < g^2 d^2$. Eventually also assumed to satisfy the inequality $d_1 d_2 - gd^2 \geq \gamma d_1 d_2$, as well as a certain condition depending on the relative size of the points z and w.
δ_1, δ_2	integers defined by $\delta_1 = \dfrac{d_1 + sd}{N}$ and $\delta_2 = \dfrac{d_2 + sd}{N}$.

Table E.1. Numbers appearing in the proof of Vojta's inequality

E.3. An Outline of the Proof of Vojta's Inequality

In this section we will give Bombieri's discourse on the ideas underlying the proof of Vojta's inequality (E.1.1). This lucid exposition is taken verbatim from Bombieri [1], Section VI; we thank Professor Bombieri for his permission to include it here.

> As early as 1965, Mumford showed that the height h_Δ on $C \times C$ could be expressed, up to bounded quantities, in terms of the Néron–Tate heights on the Jacobian of C. It then follows that
>
> $$h_\Delta(z, w) = (\text{quadratic form}) + (\text{linear form}) + O(1)$$
>
> by the quadratic nature of heights on abelian varieties. Since Δ is an effective curve on $C \times C$, the left-hand side of this equation is bounded below by a constant. On the other hand, one sees directly that the quadratic form in the right-hand side of this equation is indefinite, if the genus g is at least 2. This puts strong restrictions on the pair (z, w) because it means that z and w, considered as points in the Mordell–Weil group of the Jacobian, can never by nearly parallel with respect to the positive definite inner product determined by the Néron pairing. A simple geometric argument now shows that the heights of rational points on C, arranged in increasing order, grow at least exponentially. This is in sharp contrast with the quadratic growth one encounters on elliptic curves, and shows that rational points on curves of genus 2 or more are much harder to come by.
>
> The diagonal $\Delta = \Omega(1,1,1)$ is a Vojta divisor except for the fact that the inequalities (6) characterizing a Vojta divisor are not satisfied. However, it is easy to see that Mumford's method applies generally to any divisor that is a linear combination of $A \times C$, $C \times A$, and Δ. Now the condition $d_1 d_2 < g^2 d^2$ simply expresses the fact that the associated quadratic form is indefinite, and again we get a useful result if we can show that the height is bounded below. As in Mumford, it suffices to have an effective (i.e., positive) curve in the equivalence class of the divisor, and now the other condition $gd^2 < d_1 d_2$ for a Vojta divisor assures, by Riemann–Roch, that multiples of $\Omega(d_1, d_2, d)$ have effective representatives.
>
> The advantage in this generalization of Mumford's result is that now we have a two-parameter family of indefinite quadratic forms at our disposal, instead of just one. Thus one is tempted, given (z, w), to choose a quadratic form such that its value at (z, w) is negative, which would yield a contradiction unless z and w have bounded height.

The new problem one faces here is the fact that the choice of the quadratic form depends on the ratio of the heights of z and w, and therefore we need not only that h_Ω is bounded below, but also that the lower bound has sufficiently good uniformity with respect to the quadratic form. This is where arithmetic intersection theory and arithmetic Riemann–Roch have been used: arithmetic Riemann–Roch for finding a good effective representative for Ω defined by equations with "small" coefficients, and arithmetic intersection theory for precise control of the unwieldy bounded terms arising in the elementary theory of heights.

As Vojta's paper [1] clearly shows, this idea is overly simple and there is one more big obstacle to overcome. The argument used in obtaining a lower bound for h_Ω fails if the effective representative for Ω goes through the point (z, w) we are studying. By an appropriate use of derivations, one sees that this is not too serious a difficulty unless the representative of the divisor Ω goes through (z, w) with very high multiplicity. On the other hand, this representative must be defined by equations with small coefficients and there is very little room for moving away from (z, w), so one cannot exclude a priori that this divisor has very high multiplicity at (z, w).

This situation is reminiscent of the familiar difficulty in Diophantine approximation and transcendence theory, namely the nonvanishing at specific points of functions arising from auxiliary constructions. In the classical case, various independent techniques have been devised for this purpose: Roth's lemma, which is arithmetic in nature, the algebro-geometric Dyson's lemma, and the zero estimates of Masser and Wüstholz.

Vojta, by proving a suitable generalization of Dyson's lemma, shows that if $d_1 d_2$ is sufficiently close to gd^2, then any effective representative for $\Omega(d_1, d_2, d)$ does not vanish too much at (z, w), thereby completing the proof.

Faltings proceeds in a different way, using a new geometric tool, the product theorem. He is able to show that the difficulty with high multiplicity can be eliminated, except perhaps for a set of "bad" points (z, w) that is contained in a product subvariety of $C \times C$. It should be noted that Faltings' result applies not only to $C \times C$, but in fact to a product of an arbitrary number of varieties, thus providing a tool for handling higher-dimensional varieties by induction on the dimension.

The proof of Vojta's inequality (Theorem 2.1) thus consists of the following steps.

Step I: An Upper Bound for h_Ω (Section E.4)
We apply Mumford's method to express the height h_Ω relative to a Vojta

divisor Ω in terms of the canonical height bilinear form $\langle\cdot,\cdot\rangle$ on the Jacobian J of C. This gives us a very explicit upper bound for the height.

Step II: A Lower Bound for $h_\Omega(z,w)$ (Sections E.5, E.8, E.10)
Given a positive divisor in the divisor class of Ω, we derive a lower bound for the height $h_\Omega(z,w)$ in terms of the "size" of an equation $s = 0$ defining the positive divisor. To illustrate the general method, we first derive a lower bound under the assumption that $s(z,w) \neq 0$. Later we deal with the general case and find a lower bound that also depends on the order of vanishing of s at (z,w).

The remaining steps in the proof involve obtaining estimates for the various terms appearing in the lower bound in Step II. Eventually we want to show that the lower bound can essentially be replaced by 0, but there is a fair amount of rather technical work involved in dealing with the many terms that appear.

Step III: A Small Section (Sections E.6, E.7)
We use Siegel's lemma to produce a section for $\mathcal{O}(\Omega)$ given by a homogeneous polynomial with reasonably small coefficients. In other words, we find a "small" s to use in Step II.

Step IV: Estimating Derivatives (Section E.9)
In calculating the order of vanishing of s at (z,w), we are forced to differentiate s, so we prove an estimate (essentially due to Eisenstein) for the denominators of the derivatives of algebraic functions.

Step V: A Nonvanishing Derivative (Section E.11)
We use Roth's lemma (in fact, only the two-variable case is needed) to show that the function s from Step II does not vanish to too high an order at (z,w).

Combining the inequalities from Steps I–V, a little algebra will yield Vojta's inequality (Section E.12).

E.4. An Upper Bound for $h_\Omega(z,w)$

In this section we begin the proof of Vojta's inequality. Using the formalism of the Height Machine, we will give an upper bound for the height $h_{C\times C,\Omega}(z,w)$ of two points $z,w \in C(\bar{K})$. However, it is essential to keep track of the dependence of the error terms on d_1, d_2, d, so we will make use of the formula (7) for h_Ω in which that dependence is made explicit. The underlying idea of the following proposition is due to Mumford [1].

Proposition E.4.1. *There is a constant c_1, depending on the choice of various Weil height functions associated to the divisors A and B, such that*

for all positive integers d_1, d_2, d and all points $z, w \in C(\bar{K})$,

$$h_{C\times C,\Omega(d_1,d_2,d)}(z,w) \le \frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle + c_1(d_1+d_2+d).$$

N.B. The constant c_1 is independent of d_1, d_2, d.

PROOF. Let $j_A : C \to J$ and $p_1, p_2, s_{12} : J\times J \to J$ be the maps described in (E.2.1), and recall the definition (7)

$$\begin{aligned} h_{C\times C,\Omega(d_1,d_2,d)}(z,w) &= \delta_1 h_{C,NA}\big(p_1(z,w)\big) + \delta_2 h_{C,NA}\big(p_2(z,w)\big) \\ &\quad - d h_{C\times C,B}(z,w), \end{aligned}$$

where

$$B = Mp_1^*A + Mp_2^*A + (-\Delta + p_1^*A + p_2^*A).$$

Applying the linearity and functoriality properties of the Height Machine (B.3.2) to the height functions $h_{C,NA}$ and $h_{C\times C,B}$, we find that

$$\begin{aligned} h_{C,NA} &= Nh_{C,A} + O(1), \\ h_{C\times C,B} &= Mh_{C,A}\circ p_1 + Mh_{C,A}\circ p_2 + h_{C\times C,-\Delta+p_1^*A+p_2^*A} + O(1). \end{aligned}$$

Here the $O(1)$ depends on the choice of particular Weil height functions, but it is clearly independent of d_1, d_2, d. Substituting into the above formula for $h_{C\times C,\Omega}$ gives (note that $\delta_i = (d_i + Md)/N$)

$$\begin{aligned} h_{C\times C,\Omega(d_1,d_2,d)}(z,w) &= d_1 h_{C,A}(z) + d_2 h_{C,A}(w) \\ &\quad - d h_{C\times C,-\Delta+p_1^*A+p_2^*A}(z,w) + O(d_1+d_2+d). \end{aligned} \tag{9}$$

(Note that M is fixed, so it can be absorbed into the $O(1)$ constant.)

From (E.2.1(b)) we have $j_A^*\Theta_A \sim gA$, which gives the height relation

$$\begin{aligned} h_{C,A}(u) &= \frac{1}{g}h_{J,\Theta_A}\circ j_A(u) + O(1) \\ &= \frac{1}{g}\hat{h}_{J,\Theta}\circ j_A(u) + O(1) \\ &= \frac{1}{g}|u|^2 + O(1) \quad \text{for all } u \in C(\bar{K}). \end{aligned} \tag{10}$$

Similarly, (E.2.1(c)) says that

$$(j_A \times j_A)^*(s_{12}^*\Theta_A - p_1^*\Theta_A - p_2^*\Theta_A) \sim -\Delta + p_1^*A + p_2^*A,$$

which translates via the Height Machine into

$$\begin{aligned}
h_{C\times C,-\Delta+p_1^*A+p_2^*A}(z,w) \\
&= h_{C\times C,(j_A\times j_A)^*(s_{12}^*\Theta_A-p_1^*\Theta_A-p_2^*\Theta_A)}(z,w)+O(1) \\
&= h_{J,\Theta_A}\bigl(s_{12}(j_A(z),j_A(w))\bigr) - h_{J,\Theta_A}\bigl(p_1(j_A(z),j_A(w))\bigr) \\
&\qquad - h_{J,\Theta_A}\bigl(p_2(j_A(z),j_A(w))\bigr)+O(1) \\
&= \hat h_{J,\Theta}\bigl(j_A(z)+j_A(w)\bigr) - \hat h_{J,\Theta}\bigl(j_A(z)\bigr) - \hat h_{J,\Theta}\bigl(j_A(w)\bigr)+O(1) \\
&= |z+w|^2-|z|^2-|w|^2+O(1) \\
&= 2\langle z,w\rangle + O(1). \qquad (11)
\end{aligned}$$

Substituting (10) and (11) into (9) gives the desired result,

$$h_{C\times C,\Omega(d_1,d_2,d)}(z,w) = \frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle + O(d_1+d_2+d).$$

□

Before proceeding with the more difficult task of giving a lower bound for h_Ω, we briefly make some observations that will help to explain the significance of Proposition E.4.1 and how it is related to Vojta's inequality. First, if we set $d_1=d_2=d=1$, then $\Omega=\Omega(1,1,1)$ is just the diagonal divisor Δ, so Proposition E.4.1 becomes

$$h_{C\times C,\Delta}(z,w) \le \frac{1}{g}|z|^2 + \frac{1}{g}|w|^2 - 2\langle z,w\rangle + O(1).$$

This should look familiar; it is a slightly strengthened version of Mumford's gap principle (B.6.6(a)). (The reason we got a slightly stronger inequality than (B.6.6) is because we chose a particularly nice embedding of C in J.) In particular, since the diagonal is a positive (albeit immovable) divisor, we know from (B.3.2(c)) that $h_{C\times C,\Delta}$ is bounded below for points not lying on Δ, which implies that

$$\langle z,w\rangle \le \frac{1}{2g}\bigl(|z|^2+|w|^2\bigr)+O(1) \qquad \text{for all } z,w\in C(\bar K),\ z\ne w. \qquad (12)$$

Notice that if $g=1$, then (12) is trivially true by the Cauchy–Schwarz inequality. But if $g>1$, then (12) gives a nontrivial geometric constraint on the set $C(\bar K)$ sitting inside the Euclidean vector space $J(\bar K)\otimes\mathbb{R}$. Writing $\langle z,w\rangle = |z|\,|w|\cos\theta(z,w)$ as in the proof of Proposition E.1.2, it is easy to see what that constraint is roughly:

$$\cos\theta(z,w) \le \frac{1}{2g}\left(\frac{|z|}{|w|}+\frac{|w|}{|z|}\right).$$

Hence if z and w have approximately the same length, the right-hand side will be strictly less than 1, so $\theta(z,w)$ will be bounded away from 0. In

other words, Mumford's inequality (12) says that points in $C(\bar{K})$ of approximately the same length must subtend an angle that is bounded away from 0, and that bound is independent of the lengths. Using the argument we gave in (E.1.2) (or see Section B.6, especially Lemma B.6.7), it is a simple matter to deduce from this that the number of points in $C(K)$ of norm less than H grows no faster than $O(\log H)$. Since the number of points in $J(K)$ of norm less than H grows like a polynomial in H, we have reproven Mumford's theorem (B.6.5), which says that rational points on C are very sparsely distributed in J.

Suppose we try to apply Mumford's argument to a general Vojta divisor $\Omega = \Omega(d_1, d_2, d)$. In the first place, we need Ω to be effective, since we want h_Ω to be bounded below. As we will see later, the inequality $d_1 d_2 > gd^2$ and the Riemann–Roch theorem for surfaces will tell us that Ω is effective. If we take a particular effective divisor V in $\mathcal{O}(\Omega)$, Proposition E.4.1 gives the inequality

$$-c_2 \le h_{C\times C,V}(z,w) \le \frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle + O(d_1+d_2+d), \qquad \text{valid for } (z,w) \notin V.$$

(Of course, c_2 depends on V, so it depends on d_1, d_2, d, but let us ignore this unpleasant fact for the moment.) A little algebra now gives the estimate

$$\langle z,w\rangle \le \frac{1}{2g}\left(\frac{d_1}{d}|z|^2 + \frac{d_2}{d}|w|^2\right) + O\left(\frac{d_1}{d} + \frac{d_2}{d} + 1\right). \tag{13}$$

Taking $d_1 = d_2 = d$ retrieves Mumford's inequality, but we want to do better, so we try to make the right-hand side as small as possible, keeping in mind the constraint $d_1 d_2 > gd^2$.

Assuming that $|z|$ and $|w|$ are reasonably large (which is okay, since $C(K)$ has only finitely many points of bounded height), we might choose the integers d_1, d_2, and d to satisfy

$$d_1 \approx \sqrt{g}\,|w|^2, \qquad d_2 \approx \sqrt{g}\,|z|^2, \qquad d \approx |z|\,|w|.$$

Then (13) becomes (approximately)

$$\langle z,w\rangle \le \frac{|z|\,|w|}{\sqrt{g}} + O\left(\frac{|w|}{|z|} + \frac{|z|}{|w|}\right).$$

In particular, if $|z|$ and $|w|$ are sufficiently large, then the error term is small in comparison to the main term, and we obtain (for any $\varepsilon > 0$)

$$\langle z,w\rangle \le (1+\varepsilon)\frac{|z|\,|w|}{\sqrt{g}}.$$

Finally, the assumption that $g \geq 2$ and an appropriate choice for ε gives the nontrivial bound

$$\langle z, w\rangle \leq (1+\varepsilon)\frac{|z|\,|w|}{\sqrt{2}} \leq \frac{3}{4}|z|\,|w|,$$

which is exactly Vojta's inequality.

This "proof" was much too easy, so there must be a gap in our argument. In fact, there are two of them, both arising from the fact that we started by choosing an effective divisor V in the divisor class of $\Omega(d_1,d_2,d)$. Having chosen V, we then asserted that the height function $h_{C\times C,V}$ is bounded below. The first problem is that although this function is indeed bounded below, the $O(1)$ in the bound certainly depends on fixing particular local equations to define the divisor V. So we will need to choose equations carefully with coefficients that are not too large. This is clearly a situation where some variant of Siegel's lemma should be useful.

The second problem is more serious. We get a lower bound for $h_{C\times C,V}$ only for points (z,w) not lying on V. But as we have just seen, the particular Vojta divisor $\Omega(d_1,d_2,d)$ that we choose depends on the points (z,w), so it is at least conceivable that every effective divisor linearly equivalent to our chosen $\Omega(d_1,d_2,d)$ might contain (z,w). This problem is solved in two steps. The first step is to allow (z,w) to lie on V and to find a lower bound in terms of the multiplicity with which V goes through (z,w). The second step, which is technically more difficult, is to apply some version of Roth's lemma to show that this multiplicity is not too large; in fact, it is so small that the correction term coming from the multiplicity contributes only to the error term.

So now we have laid out the route leading to a lower bound for h_Ω. In the subsequent sections we will follow this road step by step to the desired conclusion.

E.5. A Lower Bound for $h_\Omega(z,w)$ for Nonvanishing Sections

At the end of the last section we talked about choosing an effective divisor linearly equivalent to a Vojta divisor Ω. This is the same as choosing a global section s to the line bundle $\mathcal{O}(\Omega)$, since if $s \in H^0\big(C\times C, \mathcal{O}(\Omega)\big)$ is such a section, then by definition the divisor $\operatorname{div}(s)$ is positive and linearly equivalent to Ω. We are now going to describe these global sections more explicitly.

Recall that the Vojta divisor $\Omega = \Omega(d_1,d_2,d)$ equals

$$\Omega = \delta_1(NA\times C) + \delta_2(C\times NA) - dB,$$

where δ_1, δ_2, d are assumed to be large. As discussed earlier, every global section s_1 to $\mathcal{O}(dB)$ is a homogeneous polynomial of degree d in the variables $y = [y_0, \dots, y_m]$. Further, if s is a global section to $\mathcal{O}(\Omega)$, then ss_1 is a global section to $\mathcal{O}\big(\delta_1(NA\times C)+\delta_2(C\times NA)\big)$, so ss_1 is a bihomogeneous polynomial of bidegree (δ_1, δ_2) in the variables

$$(x, x') = \big([x_0, \dots, x_n], [x'_0, \dots, x'_n]\big).$$

Applying this with s_1 taken successively to be $y_0^d, y_1^d, \dots, y_m^d$, we find that a global section s to $\mathcal{O}(\Omega)$ determines a collection of functions

$$s = \left(\frac{F_i(x,x')}{y_i^d}\right)\bigg|_{C\times C}, \tag{14}$$

where each F_i is bihomogeneous of bidegree (δ_1, δ_2) and

$$\frac{F_i}{y_i^d} = \frac{F_j}{y_j^d} \quad \text{on } C \times C, \text{ for all } 0 \le i, j \le m. \tag{15}$$

Conversely, given a set of bihomogeneous polynomials $\mathcal{F} = \{F_i(x, x')\}$ of bidegree (δ_1, δ_2) satisfying (15), the description (14) defines a global section to $\mathcal{O}(\Omega)$.

Let s be a global section of $\mathcal{O}(\Omega)$. Then the height h_Ω is bounded below for all points $(z, w) \in C \times C$ not lying on the positive divisor div(s); or equivalently, it is bounded below at all points with $s(z, w) \ne 0$. Unfortunately, the general theory of Weil heights gives us only a lower bound for h_Ω that depends on Ω, so the lower bound depends on d_1, d_2, d, and on the collection of polynomials $\mathcal{F}$ that describe s. The next proposition makes the lower bound explicit.

Proposition E.5.1. *Let s be a global section to $\mathcal{O}(\Omega)$, and let $\mathcal{F} = \{F_i\}$ be a collection of rational functions corresponding to s as above. Then for all points $(z, w) \in C \times C$ with $s(z, w) \ne 0$, we have*

$$h_{C\times C,\Omega(d_1,d_2,d)}(z, w) \ge -h(\mathcal{F}) - n \log\big((\delta_1 + n)(\delta_2 + n)\big).$$

[For notational convenience, we are writing $h(\mathcal{F}) = \max h(F_i)$, where the height of a polynomial is the height of its coefficients.]

Remark. Unfortunately, Proposition E.5.1 is not strong enough to use for the proof of Vojta's inequality, because we will not be able to guarantee the existence of a section s that does not vanish at (z, w). So later we will need to prove a stronger version in which we differentiate s until it eventually does not vanish at (z, w). The reason we give a proof of (E.5.1) is that it illustrates the general idea while keeping the technical complications to a minimum. In the next section we will show that it is possible to find a section s whose height $h(\mathcal{F})$ is fairly small, so combined with (E.5.1) we will deduce Vojta's inequality, provided that that particular s satisfies $s(z, w) \ne 0$. For the stronger version of Proposition E.5.1 for derivatives of sections, see Propositions E.8.1 and E.10.1.

PROOF (of Proposition E.5.1). By abuse of notation, we will write

$$x = \phi_{NA}(z), \qquad x' = \phi_{NA}(w), \qquad y = \phi_B(z,w).$$

Then using the particular Weil height we have associated to the Vojta divisor Ω, we have

$$\begin{aligned}
&h_{C\times C,\Omega}(z,w)\\
&\quad= \delta_1 h(x) + \delta_2 h(x') - dh(y)\\
&\quad= \delta_1 \sum_v \max_j \log|x_j|_v + \delta_2 \sum_v \max_{j'} \log|x'_{j'}|_v - d\sum_v \max_i \log|y_i|_v\\
&\quad= -\Big(\delta_1 \sum_v \min_j \log|x_j|_v^{-1}\\
&\qquad\quad + \delta_2 \sum_v \min_{j'} \log|x'_{j'}|_v^{-1} + d\sum_v \max_i \log|y_i|_v\Big)\\
&\quad= -\sum_v \max_i \min_{j,j'} \log\left|\frac{y_i^d}{x_j^{\delta_1} {x'_{j'}}^{\delta_2}}\right|_v .
\end{aligned}$$

Next we want to use our assumption that $s(z,w) \neq 0$. The standard way to exploit the fact that an algebraic number is not zero is to use the product formula, which after taking logarithms becomes

$$\sum_v \log\big|s(z,w)\big|_v = 0. \tag{16}$$

Although it has been "jazzed up" a bit, the product formula in this context is nothing more that the fact that a nonzero integer must have absolute value at least 1. Subtracting (16) from our formula for $h_\Omega(z,w)$ gives

$$h_{C\times C,\Omega}(z,w) = -\sum_v \max_i \min_{j,j'} \log\left|\frac{s(z,w)y_i^d}{x_j^{\delta_1} x_{j'}^{\delta_2}}\right|_v .$$

Now we use the fact that $s(z,w) = F_i(x,x')/y_i^d$ and that each F_i is bihomogeneous of bidegree (δ_1, δ_2) to obtain

$$\begin{aligned}
h_{C\times C,\Omega}(z,w) &= -\sum_v \max_i \min_{j,j'} \log\left|\frac{F_i(x,x')}{x_j^{\delta_1} x_{j'}^{\delta_2}}\right|_v\\
&= -\sum_v \max_i \min_{j,j'} \log\left|F_i\left(\frac{x}{x_j}, \frac{x'}{x'_{j'}}\right)\right|_v .
\end{aligned} \tag{17}$$

Remember that we are trying to find a lower bound for $h_\Omega(z,w)$. For each absolute value v we want to choose j and j' to make

$$F_i\left(\frac{x}{x_j}, \frac{x'}{x'_{j'}}\right) = F_i\left(\left[\frac{x_0}{x_j}, \dots, \frac{x_n}{x_j}\right], \left[\frac{x'_0}{x'_{j'}}, \dots, \frac{x'_n}{x'_{j'}}\right]\right)$$

small. So for each v we define j_v and j'_v to be the indices satisfying

$$|x_{j_v}|_v = \max_j |x_j|_v \qquad \text{and} \qquad \left|x'_{j'_v}\right|_v = \max_{j'} |x'_{j'}|_v,$$

which implies that

$$\left|\frac{x_j}{x_{j_v}}\right|_v \le 1 \qquad \text{and} \qquad \left|\frac{x'_{j'}}{x'_{j'_v}}\right|_v \le 1 \qquad \text{for all } j, j'.$$

Then the triangle inequality gives

$$\begin{aligned} \min_{j,j'} \log \left| F_i\left(\frac{x}{x_j}, \frac{x'}{x'_{j'}}\right)\right|_v &\le \log \left| F_i\left(\frac{x}{x_{j_v}}, \frac{x'}{x'_{j'_v}}\right)\right|_v \\ &\le \log\bigl(\nu_v(F_i) \max |\text{coefficients of } F_i|_v\bigr), \quad (18) \end{aligned}$$

where $\nu_v(F_i)$ equals 1 if v is nonarchimedean, and equals the largest possible number of nonzero terms in F_i if v is archimedean. Since the number of bihomogeneous monomials of bidegree (δ_1, δ_2) in the variables $x_0, \ldots, x_n$, $x'_0, \ldots, x'_n$ is

$$\binom{\delta_1 + n}{n}\binom{\delta_2 + n}{n} \le (\delta_1 + n)^n (\delta_2 + n)^n,$$

we have

$$\begin{aligned} &\nu_v(F_i) \le (\delta_1 + n)^n (\delta_2 + n)^n && \text{if } v \text{ is archimedean,} \\ &\nu_v(F_i) = 1 && \text{if } v \text{ is nonarchimedean.} \end{aligned}$$

Now take the maximum of (18) over all i and sum over all absolute values v. Comparing with our lower bound (17) for $h_\Omega(z, w)$, we find that

$$\begin{aligned} h_{C\times C,\Omega}(z, w) &\ge -\sum_v \max_i \log\bigl(\nu_v(F_i) \max |\text{coefficients of } F_i|_v\bigr) \\ &= -\sum_v \max_i \log |\text{coefficients of } F_i|_v - \sum_{v \text{ archimedean}} \log\bigl|\nu_v(F_i)\bigr|_v \\ &\ge -h(\mathcal{F}) - n\log\bigl((\delta_1 + n)(\delta_2 + n)\bigr). \end{aligned}$$

This is exactly what we were aiming for, which completes the proof of Proposition E.5.1. □

E.6. Constructing Sections of Small Height I: Applying Riemann–Roch

In this section and the following one we will construct a "small" global section to $\mathcal{O}\big(\Omega(d_1, d_2, d)\big)$. By small we mean that it will be given by homogeneous polynomials whose coefficients are reasonably small. The main difficulty arises from the fact that we need to do the construction uniformly with respect to d_1, d_2, d.

Recall that we have fixed embeddings

$$(\phi_{NA} \times \phi_{NA}) : C \times C \longrightarrow \mathbb{P}^n_x \times \mathbb{P}^n_{x'} \qquad \text{and} \qquad \phi_B : C \times C \longrightarrow \mathbb{P}^m_y,$$

where we have used subscripts to indicate notation for the homogeneous coordinates on the various projective spaces. Then a global section to the sheaf $\mathcal{O}\big(\Omega(d_1, d_2, d)\big)$ consists of a collection of bihomogeneous polynomials

$$\big\{F_i(x, x')\big\}_{0 \le i \le m}$$

of bidegree (δ_1, δ_2) with the property that

$$\left.\frac{F_i(x, x')}{y_i^d}\right|_{C \times C} = \left.\frac{F_j(x, x')}{y_j^d}\right|_{C \times C} \qquad \text{for all } 0 \le i, j \le m.$$

We also remind the reader that δ_1, δ_2 are related to d_1, d_2, d by the formulas

$$\delta_1 = \frac{d_1 + Md}{N}, \qquad \delta_2 = \frac{d_2 + Md}{N}.$$

Since $\phi_{NA} \times \phi_{NA}$ is an embedding, the function field of $C \times C$ is given by

$$K(C \times C) = K\left(\frac{x_1}{x_0}, \frac{x_2}{x_0}, \dots, \frac{x_n}{x_0}, \frac{x_1'}{x_0'}, \frac{x_2'}{x_0'}, \dots, \frac{x_n'}{x_0'}\right).$$

(This is a slight abuse of notation, since we really mean the restriction of the functions x_j/x_0 and x_j'/x_0' to $C \times C$, and we have implicitly used the fact that $\phi_{NA}(C)$ is not contained in the hyperplane $x_0 = 0$.) Now, y_i/y_0 is a rational function on $C \times C$, so we can write it as

$$\frac{y_i}{y_0} = \frac{P_i(x, x')}{Q_i(x, x')}$$

for some bihomogeneous polynomials $P_i, Q_i \in K[x, x']$. Substituting above, our task is as follows:

(i) Find bihomogeneous polynomials $F_0, \dots, F_m \in K[x, x']$ with bidegree (δ_1, δ_2) and satisfying the conditions

$$(P_j Q_i)^d F_i\Big|_{C \times C} = (P_i Q_j)^d F_j\Big|_{C \times C} \qquad \text{for all } 0 \le i, j \le m. \tag{19}$$

(ii) Estimate the heights of the F_i's explicitly in terms of d_1, d_2, d.

This looks like a job for some variant of Siegel's lemma, since we can write the F_i's as bihomogeneous polynomials in (x, x') and treat the coefficients as our unknowns. Then (19) puts a number of linear constraints on those coefficients, and the general philosophy associated to Siegel's lemma says that there should be a solution $\mathcal{F} = \{F_i\}$ of height

$$h(\mathcal{F}) \ll \frac{\text{dim of space of all } \{F_i\}\text{'s}}{\text{dim of space of solutions to (19)}} \times h\begin{pmatrix}\text{coefficients of the}\\ \text{linear constraints}\end{pmatrix} \qquad (20)$$

Since the P_i's and Q_i's are fixed, independent of d_1, d_2, d, the height of $(P_jQ_i)^d$ satisfies $h\bigl((P_jQ_i)^d\bigr) \ll d$. So the linear constraints have coefficients whose height is bounded (more or less) by $O(d)$.

Notice that the space of all $F(x, x')\big|_{C\times C}$ of bidegree (δ_1, δ_2) is just the space of global sections to the line bundle

$$\mathcal{O}\bigl(\delta_1(NA \times C) + \delta_2(C \times NA)\bigr).$$

We can use the Riemann–Roch theorem to estimate the dimension of this space and the dimension of the space of solutions to (19), as described in the next lemma.

Lemma E.6.1. *For all d_1 and d_2 larger than some constant depending only on C, the following two estimates are true.*

(a) $\quad \ell\bigl(\Omega(d_1, d_2, d)\bigr) \ge d_1d_2 - gd^2 - (g-1)(d_1 + d_2).$

(b) $\quad \ell\bigl(\delta_1(NA \times C) + \delta_2(C \times NA)\bigr) = (N\delta_1 - g + 1)(N\delta_2 - g + 1).$

PROOF. To ease notation, we shall write

$$A_1 = A \times C \qquad \text{and} \qquad A_2 = C \times A.$$

Thus the Vojta divisor is

$$\Omega = \Omega(d_1, d_2, d) = (d_1 - d)A_1 + (d_2 - d)A_2 + d\Delta.$$

We also observe that the canonical divisor of a product of varieties is obtained by taking the sum of the pullback of the canonical divisors on each variety (see Exercise A.2.5(b) or Hartshorne [1], Exercise II.8.3). In particular,

$$K_{C\times C} = (K_C \times C) + (C \times K_C).$$

We recall from (A.4.2.2) that $K_C \in \operatorname{Div}(C)$ is a divisor of degree $2g-2$, so in terms of intersection computations, $K_{C\times C}$ behaves like $(2g-2)(A_1 + A_2)$.

In order to calculate Ω^2 and $\Omega \cdot K_{C\times C}$, we use the following intersection table for the divisors A_1, A_2, and Δ. This table is a summary of Proposition A.4.6.4.

$\cdot$	A_1	A_2	Δ
A_1	0	1	1
A_2	1	0	1
Δ	1	1	$2-2g$

Now it is simply a matter of multiplying everything out to compute

$$\begin{aligned}\frac{1}{2}(\Omega^2 - \Omega \cdot K_{C\times C}) &= \frac{1}{2}\Omega^2 - \frac{1}{2}(\Omega \cdot (2g-2)(A_1+A_2)) \\ &= d_1d_2 - gd^2 - (g-1)(d_1+d_2).\end{aligned}$$

Further, the arithmetic genus of $C \times C$ is

$$p_a(C\times C) = g^2 - 2g;$$

see Example A.4.6.3.1 and Exercise E.5.

We now apply the Riemann–Roch theorem (A.4.6.3) to the surface $C \times C$ and the divisor $\Omega = \Omega(d_1, d_2, d)$ to obtain

$$\begin{aligned}\ell(\Omega) - s(\Omega) + \ell(K_{C\times C} - \Omega) &= \frac{1}{2}\Omega \cdot (\Omega - K_{C\times C}) + 1 + p_a(C\times C) \\ &= d_1d_2 - gd^2 - (g-1)(d_1+d_2) + (g-1)^2.\end{aligned}$$

The integer $s(\Omega)$ is nonnegative (since it is equal to the dimension of the cohomology group $H^1(C\times C, \mathcal{O}(\Omega))$), so we will obtain something slightly stronger than the desired estimate, provided that we can show that $\ell(K_{C\times C} - \Omega) = 0$.

To do this, we observe that the divisor $A_1 + A_2$ is ample on $C \times C$, and that

$$(K_{C\times C} - \Omega) \cdot (A_1 + A_2) = (4g-4) - (d_1 + d_2)$$

is strictly negative, provided that d_1 and d_2 are sufficiently large. Hence $K_{C\times C} - \Omega$ cannot be linearly equivalent to an effective divisor, which shows that $\ell(K_{C\times C} - \Omega) = 0$. This completes the proof of (a).

(b) We start with the intersection computation

$$\frac{1}{2}(\delta_1NA_1+\delta_2NA_2)\cdot(\delta_1NA_1+\delta_2NA_2-K_{C\times C}) = N^2\delta_1\delta_2 - N(g-1)(\delta_1+\delta_2).$$

Applying the Riemann–Roch theorem (A.4.6.3) to the divisor $\delta_1NA_1 + \delta_2NA_2$, we obtain

$$\begin{aligned}&\ell(\delta_1NA_1+\delta_2NA_2) - s(\delta_1NA_1+\delta_2NA_2) + \ell(K_{C\times C} - \delta_1NA_1 - \delta_2NA_2) \\ &\quad = \frac{1}{2}(\delta_1NA_1+\delta_2NA_2)\cdot(\delta_1NA_1+\delta_2NA_2-K_{C\times C}) \\ &\quad\quad + 1 + p_a(C\times C) \\ &\quad = N^2\delta_1\delta_2 - N(g-1)(\delta_1+\delta_2) + 1 + g^2 - g \\ &\quad = (N\delta_1 - g + 1)(N\delta_2 - g + 1).\end{aligned}$$

Further, we see that

$$(K_{C\times C} - \delta_1 NA_1 + \delta_2 NA_2)\cdot(A_1 + A_2) = 4g - 4 - (\delta_1 N + \delta_2 N)$$

is negative for sufficiently large d_1, d_2, d, so just as in (a), we find that $\ell(K_{C\times C} - \delta_1 NA_1 - \delta_2 NA_2) = 0$. It remains to deal with the pesky "superabundance" term $s(\delta_1 NA_1 + \delta_2 NA_2)$. More precisely, we will have completed the proof of (b) once we show that the superabundance is 0.

The quickest way to show this is to invoke the powerful vanishing theorem of Kodaira (Remark 4.6.3.2). (See Hartshorne [1], Remark III.7.15 for the general statement). Kodaira's vanishing theorem tells us that if D is an ample divisor on a surface X, then $s(K_X + D) = 0$. In the present instance

$$K_{C\times C} \sim (2g-2)A_1 + (2g-2)A_2;$$

hence

$$\delta_1 NA_1 + \delta_2 NA_2 = K_{C\times C} + \underbrace{(\delta_1 N + 2 - 2g)A_1 + (\delta_2 N + 2 - 2g)A_2}_{\text{ample if } \delta_1 N > 2g-2 \text{ and } \delta_2 N > 2g-2}.$$

This proves that if $\delta_1 N$ and $\delta_2 N$ are sufficiently large, then

$$\delta_1 NA_1 + \delta_2 NA_2 = K_{C\times C} + D$$

for an ample divisor D, so Kodaira's theorem gives us the desired conclusion $s(\delta_1 NA_1 + \delta_2 NA_2) = 0$. □

Remark E.6.2. Using a variant of the Enriques–Severi–Zariski lemma, it is possible to show directly that the natural map

$$L(\delta_1 NA) \otimes L(\delta_2 NA) \longrightarrow L(\delta_1 NA_1 + \delta_2 NA_2)$$

is surjective for sufficiently large δ_1, δ_2. It follows that

$$\ell(\delta_1 NA_1 + \delta_2 NA_2) = \ell(\delta_1 NA)\ell(\delta_1 NA) = (N\delta_1 - g + 1)(N\delta_2 - g + 1).$$

Using Lemma E.6.1 in the estimate (20), we find that there should be a global section s to $\mathcal{O}(\Omega(d_1, d_2, d))$ given by a system of polynomials $\mathcal{F}$ satisfying

$$h(\mathcal{F}) \ll \frac{(d_1 + Md)(d_2 + Md)}{d_1 d_2 - gd^2}\cdot d,$$

at least provided that d_1, d_2, d are chosen sufficiently large. In the next section we will make this plausibility argument precise. The main difficulty that arises is that of translating (19) into an explicit set of linear equations while maintaining sufficient control over both the number of variables and the coefficients of those equations.

E.7. Constructing Sections of Small Height II: Applying Siegel's Lemma

In this section we combine the Riemann–Roch calculations from the previous section with some messy polynomial calculations and an application of Siegel's lemma to produce a "small" global section to $\mathcal{O}(\Omega(d_1, d_2, d))$.

Proposition E.7.1. *Let $\gamma > 0$ be given, and let d_1, d_2, d be large integers satisfying*

$$d_1 d_2 - g d^2 \geq \gamma d_1 d_2.$$

Then there is a global section s to $\mathcal{O}(\Omega(d_1, d_2, d))$ given by a system of bihomogeneous polynomials $\mathcal{F} = (F_0, \ldots, F_m)$ as described in Section E.5 such that

$$h(\mathcal{F}) \leq c_1 \frac{d_1 + d_2}{\gamma} + o(d_1 + d_2).$$

Here c_1 depends on C/K and the embeddings ϕ_{NA} and ϕ_B, but is independent of d_1, d_2, and d.

PROOF. We begin by choosing a suitable affine coordinate system with which to work. Consider the projection map

$$\begin{array}{rccc} \pi : C & \xrightarrow{\phi_{NA}} \mathbb{P}^n & \longrightarrow & \mathbb{P}^1, \\ & x & \longmapsto & [x_0, x_1]. \end{array}$$

The embedding ϕ_{NA} corresponds to choosing a basis for the linear system $L(NA)$, so the map $\pi : C \to \mathbb{P}^1$ is a nonconstant rational map. (If it were undefined or constant, then $\phi_{NA}(C)$ would lie in a hyperplane, contradicting the linear independence of the functions used to define ϕ_{NA}.) Further, since C is a smooth curve, π is automatically a morphism. It is easy to compute the degree of π:

$$\begin{aligned} \deg \pi &= \deg \pi^*(\text{a point in } \mathbb{P}^1) \\ &= \deg \phi_{NA}^*(\text{a hyperplane in } \mathbb{P}^n) = \deg NA = N. \end{aligned}$$

The third equality is true because ϕ_{NA}^* applied to any hyperplane in $\mathbb{P}^n$ gives a divisor linearly equivalent to NA.

Similarly, for each $2 \leq j \leq n$ we can project

$$\begin{array}{rccc} \pi_j : C & \xrightarrow{\phi_{NA}} \mathbb{P}^n & \longrightarrow & \mathbb{P}^2, \\ & x & \longmapsto & [x_0, x_1, x_j]. \end{array}$$

We use here the content of Remark E.2.2, which says that if the coordinates $(x_0, \ldots, x_n)$ have been nicely chosen, then the image of π_j is a (possibly singular) curve in $\mathbb{P}^2$ of degree

$$\begin{aligned} \deg \pi_j(C) &= \deg \pi_j^*(\text{line in } \mathbb{P}^2) \\ &= \deg \phi_{NA}^*(\text{hyperplane in } \mathbb{P}^n) = \deg NA = N. \end{aligned}$$

Thus $\pi_j(C)$ is a curve of degree N in $\mathbb{P}^2$. If we restrict the projection

$$\mathbb{P}^2 \longrightarrow \mathbb{P}^1, \qquad [x_0, x_1, x_j] \longmapsto [x_0, x_1],$$

to $\pi_j(C)$, we get a map of degree N from $\pi_j(C)$ to $\mathbb{P}^1$, and then the following diagram commutes.

$$\begin{array}{ccccc} C & & \xrightarrow{\pi_j} & & \pi_j(C) \subset \mathbb{P}^2 \\ & \pi \searrow \deg N & & \swarrow \deg N & \\ & & \mathbb{P}^1 & & \end{array}$$

It follows that the map $\pi_j : C \to \pi_j(C)$ has degree 1, so $\pi_j(C)$ is birational to C. In particular, the function field of C equals

$$K(C) = K\left(\left.\frac{x_1}{x_0}\right|_C, \left.\frac{x_j}{x_0}\right|_C\right) \quad \text{for any } 2 \le j \le n.$$

This suggests that we define affine coordinates on C by letting

$$\xi_j = \left.\frac{x_j}{x_0}\right|_C, \qquad 1 \le j \le n.$$

Then $K[\xi_1, \dots, \xi_n]$ is the affine coordinate ring of

$$C \smallsetminus A = C \cap \{x_0 \ne 0\}.$$

The fact that $\pi_j(C)$ is an irreducible curve of degree N in $\mathbb{P}^2$ means that it is the set of zeros of a homogeneous polynomial of degree N in the variables x_0, x_1, x_j. Without loss of generality, we may assume that this polynomial has an x_j^N term. (If not, we need merely go back and use a slightly modified basis for $L(NA)$ to define the embedding ϕ_{NA}.) Dividing the equation for $\pi_j(C)$ by cx_0^N gives a relation of the form

$$\xi_j^N = \sum_{i=0}^{N-1} a_{ij}(\xi_1)\xi_j^i, \tag{21}$$

where $a_{ij} \in K[\xi_1]$ are polynomials satisfying

$$\deg_{\xi_1} a_{ij}(\xi_1) \le N - i.$$

It will be convenient later if the a_{ij}'s have coefficients in the ring of integers R of K. We can ensure that this is true by again changing the embedding ϕ_{NA}. To be precise, if the original ϕ_{NA} was given by $[x_0, \dots, x_n]$, let the new ϕ_{NA} be $[x_0, rx_1, \dots, rx_n]$, where $r \in K^*$ is chosen to clear the denominators of all of the coefficients of all of the a_{ij}'s. Then (21) becomes

$$\xi_j^N = \sum_{i=0}^{N-1} a_{ij}(\xi_1)\xi_j^i, \qquad a_{ij} \in R[\xi_1], \qquad \deg a_{ij} \le N - i. \tag{22}$$

We know that

$$\begin{array}{ccccc} C & \overset{\deg 1}{\longrightarrow} & \pi_2(C) & \overset{\deg N}{\longrightarrow} & \mathbb{P}^1, \\ z & \longmapsto & \big[1,\xi_1(z),\xi_2(z)\big] & \longmapsto & \big[1,\xi_1(z)\big]. \end{array}$$

It follows that

$$K(C)=K(\xi_1,\xi_2) \qquad \text{and} \qquad \big[K(\xi_1,\xi_2):K(\xi_1)\big]=N.$$

In particular, $1,\xi_2,\xi_2^2,\dots,\xi_2^{N-1}$ form a basis for $K(C)$ over $K(\xi_1)$. Similarly, if we take two copies of this map,

$$C\times C \xrightarrow{([1,\xi_1,\xi_2],[1,\xi_1',\xi_2'])} \mathbb{P}^2\times\mathbb{P}^2,$$

we find that

$$K(C\times C)=K(\xi_1,\xi_2,\xi_1',\xi_2'). \tag{23}$$

Further, the set of functions

$$\mathcal{B}=\{\xi_2^i\xi_2'^j \mid 0\le i,j\le N-1\} \tag{24}$$

is a basis for $K(C\times C)$ over $K(\xi_1,\xi_1')$.

Next we use the embedding $\phi_B : C\times C\to\mathbb{P}^m$ corresponding to the divisor B and pull back the coordinate functions on $\mathbb{P}^m$. This gives rational functions $\phi_B^*(y_i/y_0)$ on $C\times C$, so (23) tells us that this function is in $K(\xi_1,\xi_2,\xi_1',\xi_2')$. (We know that the image $\phi_B(C\times C)$ does not lie in the hyperplane $y_0=0$.) Thus

$$\phi_B^*\left(\frac{y_i}{y_0}\right)=\frac{P_i(\xi_1,\xi_2,\xi_1',\xi_2')}{Q_i(\xi_1,\xi_2,\xi_1',\xi_2')}$$

for some polynomials $P_i,Q_i\in K[\xi_1,\xi_2,\xi_1',\xi_2']$, and since we are allowed to multiply numerator and denominator by a common element of K^*, we may clear denominators and assume that $P_i,Q_i\in R[\xi_1,\xi_2,\xi_1',\xi_2']$. It is important to observe that the P_i's and Q_i's are independent of d_1, d_2, and d.

Recall that we are looking for a collection of bihomogeneous polynomials $\{F_i\}$ satisfying the condition (19), since such a collection of polynomials corresponds to a global section to $\mathcal{O}\big(\Omega(d_1,d_2,d)\big)$. Dehomogenizing (19), we can rewrite the necessary conditions in terms of our affine coordinates $\xi=(\xi_1,\dots,\xi_n)$ and $\xi'=(\xi_1',\dots,\xi_n')$ as follows: A global section to $\mathcal{O}\big(\Omega(d_1,d_2,d)\big)$ corresponds to a collection of functions $F_i(\xi,\xi')\in K[\xi,\xi']$ of degree at most δ_1 in ξ and at most δ_2 in ξ', satisfying the conditions

$$(P_jQ_i)^dF_i=(P_iQ_j)^dF_j \quad \text{for all } 0\le i,j\le m. \tag{25}$$

In order to find a "small" solution to the system of equations (25), we will look at the following three vector spaces of functions:

$$V_1 = \left\{ F \in K(C \times C) \;\middle|\; \begin{array}{l} F \in K[\xi, \xi'] \text{ has degree at most } \delta_1 \text{ in } \xi \\ \text{and degree at most } \delta_2 \text{ in } \xi' \end{array} \right\},$$

$$V_2 = V_1 \cap K[\xi_1, \xi_2, \xi_1', \xi_2'],$$

$$V_3 = \big\{ (F_0, \dots, F_m) \in V_1^{m+1} \;\big|\; (P_j Q_i)^d F_i = (P_i Q_j)^d F_j \quad \text{for all } 0 \le i, j \le m \big\}.$$

Notice that if we rehomogenize ξ and ξ', then V_1 is nothing more than the space of global sections to $\mathcal{O}\big(\delta_1(NA \times C) + \delta_2(C \times NA)\big)$, so from Lemma E.6.1(b) we have

$$\dim V_1 = (N\delta_1 - g + 1)(N\delta_2 - g + 1).$$

Similarly, V_2 contains the subspace of $K(C \times C)$ spanned by the monomials

$$\xi_1^i \xi_2^j \xi_1'^{i'} \xi_2'^{j'} \quad \text{with } 0 \le j, j' \le N - 1, \quad 0 \le i + j \le \delta_1, \quad 0 \le i' + j' \le \delta_2;$$

and from (24) we see that these monomials are K-linearly independent in $K(C \times C)$ and span V_2. It is an easy matter to count the number of such monomials, which yields

$$\dim V_2 = \left(N\delta_1 - \tfrac{1}{2}N(N-3)\right)\left(N\delta_2 - \tfrac{1}{2}N(N-3)\right).$$

Notice in particular that V_2 is almost as large as V_1.

Finally, we observe that the elements in V_3 are exactly the global sections of $\mathcal{O}\big(\Omega(d_1, d_2, d)\big)$, so Lemma E.6.1 (a) gives

$$\dim V_3 \ge d_1 d_2 - g d^2 - (g-1)(d_1 + d_2).$$

Remember that we are looking in V_1^{m+1} for a small element of V_3. It will be much easier to look for that element in V_2^{m+1}, rather than in V_1^{m+1}, since we have an explicit basis for V_2. To see that V_2^{m+1} is large enough, we estimate

$$\begin{aligned} \dim V_3 \cap V_2^{m+1} &\ge \dim V_3 - (\dim V_1^{m+1} - \dim V_2^{m+1}) \\ &\ge \left\{ d_1 d_2 - g d^2 - (g-1)(d_1 + d_2) \right\} \\ &\quad - \left\{ (m+1)(N\delta_1 - g + 1)(N\delta_2 - g + 1) \right\} \\ &\quad + \left\{ (m+1)\left(N\delta_1 - \tfrac{1}{2}N(N-3)\right)\left(N\delta_2 - \tfrac{1}{2}N(N-3)\right) \right\} \\ &\ge d_1 d_2 - g d^2 + O(d_1 + d_2) \\ &\ge \gamma d_1 d_2 + O(d_1 + d_2). \end{aligned} \tag{26}$$

It remains to find explicit linear constraints on the elements of V_2^{m+1} that guarantee that they lie in V_3.

Applying (22) with $j = 2$ to both copies of $C \times C$ embedded in $\mathbb{P}^n \times \mathbb{P}^n$ via $\phi_{NA} \times \phi_{NA}$, we find that

$$\xi_2^N = \sum_{i=0}^{N-1} a_{2j}(\xi_1)\xi_2^i \qquad \text{and} \qquad {\xi_2'}^N = \sum_{i=0}^{N-1} a_{2j}(\xi_1'){\xi_2'}^i. \tag{27}$$

Now we consider two elements $\mu, \nu \in \mathcal{B}$, where $\mathcal{B}$ is our fixed basis (24) for $K(C \times C)$ over $K(\xi_1, \xi_1')$. If we form the product $\mu\nu$, then we can use (27) repeatedly to express $\mu\nu$ as an $R[\xi_1, \xi_1']$-linear combination of elements of $\mathcal{B}$. In other words, for all $\mu, \nu \in \mathcal{B}$ we have

$$\mu\nu = \sum_{\lambda \in \mathcal{B}} b_{\mu\nu\lambda}(\xi_1, \xi_1')\lambda \qquad \text{with } b_{\mu\nu\lambda} \in R[\xi_1, \xi_1']. \tag{28}$$

Next suppose that we are given a function

$$P(\xi_1, \xi_2, \xi_1', \xi_2') \in R[\xi_1, \xi_2, \xi_1', \xi_2'] \subset K(C \times C).$$

Using (27) and the fact that $\mathcal{B}$ is a basis, we can write such a P uniquely as

$$P(\xi_1, \xi_2, \xi_1', \xi_2') = \sum_{\mu \in \mathcal{B}} p_\mu(\xi_1, \xi_1')\mu \qquad \text{with } p_\mu \in R[\xi_1, \xi_1']. \tag{29}$$

Similarly, any power P^d of P can be written uniquely as

$$P(\xi_1, \xi_2, \xi_1', \xi_2')^d = \sum_{\mu \in \mathcal{B}} p_{\mu d}(\xi_1, \xi_1')\mu \qquad \text{with } p_{\mu d} \in R[\xi_1, \xi_1']. \tag{30}$$

The following lemma gives an estimate for the height of the $p_{\mu d}$'s explicitly in terms of d.

Lemma E.7.2. *Let P be as in (29) and define $p_{\mu d} \in R[\xi_1, \xi_1']$ by (30). Let $b_{\mu\nu\lambda}$ be the polynomials defined in (28).*
(a) *For all $d \geq 1$ and all $\mu \in \mathcal{B}$,*

$$p_{\mu,d+1} = \sum_{\lambda,\nu \in \mathcal{B}} p_{\lambda d} \cdot p_\nu \cdot b_{\lambda\nu\mu}.$$

(b) *There is a constant c_2, depending only on K, N, the a_{2j}'s from (27), and the $b_{\lambda\nu\mu}$'s from (28), such that for all $d \geq 1$,*

$$\max_{\mu \in \mathcal{B}} \deg p_{\mu d} \leq c_2 d \qquad \text{and} \qquad \max_{\mu \in \mathcal{B}} h(p_{\mu d}) \leq c_2 d.$$

N.B. The important point in this lemma is to keep track of the dependence on d.

PROOF (a)
$$\begin{aligned} P^{d+1} &= P^d \cdot P \\ &= \Big(\sum_{\mu\in\mathcal{B}} p_{\mu d}\mu\Big)\Big(\sum_{\nu\in\mathcal{B}} p_\nu \nu\Big) \quad \text{from (29) and (30)} \\ &= \sum_{\mu,\nu\in\mathcal{B}} \Big(p_{\mu d} p_\nu \sum_{\lambda\in\mathcal{B}} b_{\mu\nu\lambda}\lambda\Big) \quad \text{from (28)} \\ &= \sum_{\lambda\in\mathcal{B}} \Big(\sum_{\mu,\nu\in\mathcal{B}} p_{\mu d} p_\nu b_{\mu\nu\lambda}\Big)\lambda. \end{aligned}$$

Comparing this with
$$P^{d+1} = \sum_{\lambda\in\mathcal{B}} p_{\lambda,d+1}\lambda$$
gives (a), since the elements of $\mathcal{B}$ are linearly independent over $K(\xi_1,\xi_1')$.
(b) Applying (a) repeatedly, we see that $p_{\mu,d+1}$ can be expressed as a sum
$$p_{\mu,d+1} = \sum_{\lambda_1,\nu_1\in\mathcal{B}} \sum_{\lambda_2,\nu_2\in\mathcal{B}} \cdots \sum_{\lambda_d,\nu_d\in\mathcal{B}} p_{\lambda_d} p_{\nu_1} p_{\nu_2} \cdots p_{\nu_d} b_{\lambda_1\nu_1\mu} b_{\lambda_2\nu_2\lambda_1} \cdots b_{\lambda_d\nu_d\lambda_{d-1}}.$$

This shows immediately that
$$\deg p_{\mu,d+1} \le (d+1)\left(\max_{\mu\in\mathcal{B}} \deg p_\mu\right) + d\left(\max_{\lambda,\nu,\mu\in\mathcal{B}} \deg b_{\lambda\nu\mu}\right) \le c_3 d,$$
which proves half of (b).

For the other half, we note that $\#\mathcal{B} = N^2$, so we have expressed $p_{\mu,d+1}$ as a sum of N^{2d} terms, each of which has the form
$$\big(\text{product of } (d+1) \text{ of the } p_\mu\text{'s}\big) \times \big(\text{product of } d \text{ of the } b_{\mu\nu\lambda}\text{'s}\big).$$

Applying Proposition B.7.2 then yields
$$\begin{aligned} h(p_{m,d+1}) &\le [K:\mathbb{Q}] \sup h\Big(\prod_{d+1 \text{ factors}} p_\mu \times \prod_{d \text{ factors}} b_{\mu\nu\lambda}\Big) + \log N^{2d} \\ &\le [K:\mathbb{Q}]\Big((d+1)\sup_{\mu\in\mathcal{B}}\big\{h(p_\mu) + \deg p_\mu + 1\big\} \\ &\qquad + d \sum_{\mu,\nu,\lambda\in\mathcal{B}} \big\{h(b_{\mu\nu\lambda}) + \deg b_{\mu\nu\lambda} + 1\big\}\Big) + 2d\log N \\ &\le c_4(d+1). \end{aligned}$$

This completes the proof of Lemma E.7.2. □

We return now to our fixed collection of polynomials

$$P_i, Q_i \in R[\xi_1, \xi_2, \xi_1', \xi_2'].$$

We apply Lemma E.7.2 with $P = P_iQ_j$ for each $0 \le i, j \le m$, which allows us to write

$$(P_iQ_j)^d = \sum_{\mu \in \mathcal{B}} p_{ij\mu d}\mu \qquad \text{with } p_{ij\mu d} \in R[\xi_1, \xi_1'] \text{ satisfying } h(p_{ij\mu d}) \le c_5 d. \tag{31}$$

We can choose one constant c_5 to work for all of the P_iQ_j's; note that c_5 is independent of d.

Every element F_i in our space V_2 can be written uniquely in the form

$$F_i = \sum_{\substack{\nu \in \mathcal{B} \\ k,k' \ge 0}} u_{ikk'\nu}\xi_1^k \xi_1'^{k'} \nu \in V_2. \tag{32}$$

Now, an $(m+1)$-tuple $(F_0, \dots, F_m) \in V_2^{m+1}$ will lie in V_3 if and only if it satisfies the conditions

$$0 = (P_jQ_i)^d F_i - (P_iQ_j)^d F_j \qquad \text{for all } 0 \le i, j \le m. \tag{33}$$

Using (31) and (32) to expand (33), we find that (33) is equivalent to

$$\begin{aligned}
0 &= \Big(\sum_{\mu \in \mathcal{B}} p_{ji\mu d}\mu\Big)\Big(\sum_{\substack{\nu \in \mathcal{B} \\ k,k' \ge 0}} u_{ikk'\nu}\xi_1^k \xi_1'^{k'} \nu\Big) \\
&\quad - \Big(\sum_{\mu \in \mathcal{B}} p_{ij\mu d}\mu\Big)\Big(\sum_{\substack{\nu \in \mathcal{B} \\ k,k' \ge 0}} u_{jkk'\nu}\xi_1^k \xi_1'^{k'} \nu\Big) \\
&= \sum_{\substack{\mu,\nu \in \mathcal{B} \\ k,k' \ge 0}} \Big(\xi_1^k \xi_1'^{k'} (p_{ji\mu d}u_{ikk'\nu} - p_{ij\mu d}u_{jkk'\nu}) \sum_{\lambda \in \mathcal{B}} b_{\mu\nu\lambda}\lambda\Big) \quad \text{from (28)} \\
&= \sum_{\lambda \in \mathcal{B}} \Big\{ \sum_{\substack{\mu,\nu \in \mathcal{B} \\ k,k' \ge 0}} \xi_1^k \xi_1'^{k'} b_{\mu\nu\lambda}(p_{ji\mu d}u_{ikk'\nu} - p_{ij\mu d}u_{jkk'\nu})\Big\}\lambda.
\end{aligned} \tag{34}$$

Keep in mind that our goal is to find $u_{ikk'\nu}$ coefficients such that (34) is true for all $0 \le i, j \le m$. So the $u_{ikk'\nu}$'s are our variables, and (34) gives us a system of linear constraints on these variables. Further, our calculation of the dimension of $V_3 \cap V_2^{m+1}$ tells us that (33), and hence also (34), has a solution space of dimension at least $\gamma d_1 d_2 + O(d_1 + d_2)$.

Notice that the braced expression in the right-hand side of (34) is a polynomial in ξ_1 and ξ_1' only; it is free of ξ_2's and ξ_2''s. Since the elements

of $\mathcal{B}$ are linearly independent over $K(\xi_1, \xi_1')$, it follows that (34) is true for all $0 \le i, j \le m$ if and only if

$$\sum_{\substack{\mu,\nu\in\mathcal{B}\\ k,k'\ge 0}} \xi_1^k {\xi_1'}^{k'} b_{\mu\nu\lambda}(p_{ji\mu d}u_{ikk'\nu} - p_{ij\mu d}u_{jkk'\nu}) = 0 \qquad (35)$$
$$\text{for all } 0 \le i, j \le m \text{ and } \lambda \in \mathcal{B}.$$

The fact that (34) and (35) are equivalent is important, because (26) tells us that the space of solutions to (34) has dimension at least $\gamma d_1 d_2 + O(d_1 + d_2)$, so we deduce that

$$\dim\bigl(\text{space of solutions to (35)}\bigr) \ge \gamma d_1 d_2 + O(d_1 + d_2). \qquad (36)$$

On the other hand, the functions in (35) lie in the ring $R[\xi_1, \xi_1'] \subset K(C \times C)$, and ξ_1 and ξ_1' are algebraically independent over K. In other words, $R[\xi_1, \xi_1']$ is just a polynomial ring, so (35) will be true if and only if the coefficient of each distinct monomial $\xi_1^\ell {\xi_1'}^{\ell'}$ is equal to zero. So it remains to rewrite (35) as a sum of monomials and set the coefficients equal to zero, thereby obtaining the linear constraints that the $u_{ikk'\nu}$'s must satisfy if our system $\mathcal{F} = (F_0, \ldots, F_m)$ is to be a global section of $\mathcal{O}(\Omega)$.

We know that the $b_{\mu\nu\lambda}(\xi_1, \xi_1')$'s and the $p_{ij\mu d}(\xi_1, \xi_1')$'s are polynomials whose degrees and heights satisfy

$$\deg b_{\mu\nu\lambda} \le c_6, \quad h(b_{\mu\nu\lambda}) \le c_7, \qquad \deg p_{ij\mu d} \le c_8 d, \quad h(p_{ij\mu d}) \le c_9 d.$$

(The first two inequalities are clear, since the $b_{\mu\nu\lambda}$'s do not depend on d, while the second two inequalities are Lemma E.7.2(b).) Hence if we let

$$f_{kk'\mu\nu\lambda ijd}(\xi_1, \xi_1') = \xi_1^k {\xi_1'}^{k'} b_{\mu\nu\lambda}(\xi_1, \xi_1') p_{ij\mu d}(\xi_1, \xi_1'),$$

then

$$h(f_{kk'\mu\nu\lambda ijd}) \le c_{10} d.$$

Now our system of linear constraints becomes

$$\sum_{\substack{\mu,\nu\in\mathcal{B}\\ k,k'\ge 0}} \bigl(f_{kk'\mu\nu\lambda jid}(\xi_1, \xi_1') u_{ikk'\nu} - f_{kk'\mu\nu\lambda ijd}(\xi_1, \xi_1') u_{jkk'\nu}\bigr) = 0$$
$$\text{for all } 0 \le i, j \le m \text{ and } \lambda \in \mathcal{B}.$$

Setting the coefficient of each distinct monomial $\xi_1^\ell {\xi_1'}^{\ell'}$ equal to zero, we see that the linear constraints on the $u_{ikk'\nu}$'s have coefficients bounded by $c_{11} d$. This estimate is what we have been aiming for.

In summary, we have shown the following:

(i) $\dim V_3 \cap V_2^{m+1} \ge \gamma d_1 d_2 + O(d_1 + d_2)$.

(ii) $\dim V_2^{m+1} \le (m+1)N^2\delta_1\delta_2 + O(d_1+d_2+d)$.
This follows from Lemma E.5.1(b), since V_2 is contained in V_1, and Lemma E.5.1(b) gives such an upper bound for $\dim V_1$.

(iii) Elements of $V_3 \cap V_2^{m+1}$ are precisely those $(m+1)$-tuples of functions $\mathcal{F} = (F_0, \dots, F_m)$, where each F_i is given by (32) and the coefficients $u_{ikk'\nu}$ in (32) are constrained to satisfy a system of homogeneous linear equations whose coefficients have height bounded by $c_{11}d$.

[Aside: Note that (i) and (ii) certainly lie much deeper than (iii), since they depend on the Riemann–Roch theorem. The proof of (iii) was long, but essentially nothing more than an elementary calculation.]

Applying Siegel's lemma (Proposition D.4.2), we find that there is an element

$$\mathcal{F} = (F_0, \dots, F_m) \in V_3 \cap V_2^{m+1}$$

satisfying

$$\begin{aligned} h(\mathcal{F}) &\le c_{12}\frac{\dim V_2^{m+1}}{\dim V_3 \cap V_2^{m+1}} \cdot h(\text{coefficients of linear constraints}) \\ &\le c_{13}\frac{(m+1)N^2\delta_1\delta_2 + O(d_1+d_2+d)}{\gamma d_1 d_2 + O(d_1+d_2)} \cdot d. \end{aligned}$$

Now using the definition $\delta_i = (d_i + Md)/N$ and the assumption that d_1, d_2, d are large and satisfy $d_1 d_2 > g d^2$, a little algebra yields the desired result,

$$h(\mathcal{F}) \le c_{14}\frac{d_1+d_2}{\gamma} + o(d_1+d_2).$$

This completes the proof of Proposition E.7.1. □

E.8. Lower Bound for $h_\Omega(z,w)$ at Admissible (i_1^*, i_2^*): Version I

Recall that in Section E.5 we proved a lower bound for $h_\Omega(z,w)$ of the form

$$h_{C\times C,\Omega(d_1,d_2,d)}(z,w) \ge -h(\mathcal{F}) - n\log\big((\delta_1+n)(\delta_2+n)\big), \qquad (37)$$

where $\mathcal{F} = \{F_i\}$ is a collection of bihomogeneous polynomials describing the global section s of $\mathcal{O}(\Omega)$. Further, we now know from Section E.7 that there is such a section whose height is bounded by

$$h(\mathcal{F}) \le c_{15}\frac{d_1+d_2}{\gamma} + o(d_1+d_2),$$

so we can refine the lower bound (37) to

$$h_{C\times C,\Omega(d_1,d_2,d)}(z,w) \geq -c_{15}\frac{d_1+d_2}{\gamma} + o(d_1+d_2) - n\log\bigl((\delta_1+n)(\delta_2+n)\bigr).$$

This last lower bound, combined with the argument sketched at the end of Section E.4, suffices to prove Vojta's inequality. So what is still left to do?

The one remaining problem is that (37) was proven only under the additional assumption that the section s does not vanish at (z, w), and unfortunately in order to prove Vojta's inequality, we need to choose d_1, d_2, d depending on z and w. How do we know that the corresponding section described in Proposition E.7.1 does not vanish at (z, w)? The answer is that we do not, it is quite likely that in fact $s(z, w)$ is equal to 0. So taking our cue from the classical Thue–Siegel–Roth proof, we start differentiating s until we find some derivative that does not vanish at (z, w). We then redo the entire argument to obtain an unconditional lower bound for $h_\Omega(z, w)$.

The task we have before us divides naturally into three steps. In this section we will assume that some derivative of s does not vanish at (z, w) and derive a lower bound for $h_\Omega(z, w)$ depending on the order of the derivative and the values of the derivatives of certain algebraic functions. In the next section we will give an estimate for those derivatives that depends in a very explicit way on the order of the derivative and on z and w. All of these estimates are elementary, although notationally they are somewhat intricate. Then, in Section E.10, we will show that it is not necessary to differentiate s too many times in order to obtain a nonvanishing derivative. In order to do this we will apply a two-variable version of Roth's lemma that was essentially already used by Siegel.

We begin with some notation. We let ζ and ζ' be uniformizers at the points z and w, respectively. Then any rational function on $C \times C$ that is regular at (z, w) can be thought of as a function of ζ and ζ', so we can compute its partial derivatives with respect to ζ and ζ'. We define differential operators

$$\partial_i = \frac{1}{i!}\left(\frac{\partial}{\partial\zeta}\right)^i, \qquad \partial'_i = \frac{1}{i!}\left(\frac{\partial}{\partial\zeta'}\right)^i. \tag{38}$$

We will assume throughout that z and w satisfy

$$\begin{aligned} x_j(z) &\neq 0 \quad \text{for all } 0 \leq j \leq n,\\ x'_{j'}(w) &\neq 0 \quad \text{for all } 0 \leq j' \leq n,\\ y_0(z,w) &\neq 0. \end{aligned} \tag{39}$$

In other words, we are discarding a finite number of points of C where it intersects certain hyperplanes for the embeddings ϕ_{NA}. Remember that by construction, C is not contained in any of these hyperplanes. The last condition is clearly satisfied, perhaps after changing the order of the y_i's.

Let s be a global section of $\mathcal{O}(\Omega(d_1,d_2,d))$ corresponding to a collection of polynomials $\{F_i\}$ as usual. We define a rational function $f \in K(C\times C)$ by

$$f = \frac{s}{x_0^{\delta_1} {x'_0}^{\delta_2} y_0^{-d}} = \left(\frac{y_0}{y_i}\right)^d F_i\left(\frac{x}{x_0}, \frac{x'}{x'_0}\right).$$

Here we are writing x/x_0 for the $(n+1)$-tuple $(1, x_1/x_0, \dots, x_n/x_0)$, and similarly for x'/x'_0. We are interested in studying some (hopefully small) derivative of f that does not vanish at (z,w). With this in mind, we make the following definitions.

Definition. The *index of the section s at the point* (z,w) is

$$\operatorname{Ind}(s) = \min\left\{ \frac{i_1}{\delta_1} + \frac{i_2}{\delta_2} \,\middle|\, i_1, i_2 \ge 0 \quad\text{and}\quad \partial_{i_1}\partial'_{i_2} f(z,w) \ne 0 \right\}.$$

A pair (i_1^*, i_2^*) is called *admissible for s* if

$$\operatorname{Ind}(s) = \frac{i_1^*}{\delta_1} + \frac{i_2^*}{\delta_2} \qquad\text{and}\qquad \partial_{i_1^*}\partial'_{i_2^*} f(z,w) \ne 0.$$

Thus the index of s measures the smallest derivative of s that does not vanish at (z,w), and an admissible pair gives a particular derivative realizing this minimum. We start with a simple application of Leibniz's rule.

Lemma E.8.1. *With notation as above, let (i_1^*, i_2^*) be an admissible pair for s. Let g be a rational function on $C\times C$ that is regular and nonvanishing at (z,w). Then*

$$\left(\partial_{i_1^*}\partial'_{i_2^*} f\right)(z,w) = \left(\frac{\partial_{i_1^*}\partial'_{i_2^*}(fg)}{g}\right)(z,w).$$

PROOF. Leibniz's rule for the derivative of a product says that

$$\partial_{i_1^*}\partial'_{i_2^*}(fg) = \sum_{u+v=i_1^*}\ \sum_{u'+v'=i_2^*} (\partial_u \partial'_{u'} f)(\partial_v \partial'_{v'} g). \tag{40}$$

Note that the definition (38) of ∂ and ∂' includes factorials that take care of the usual combinatorial quantities appearing in Leibniz's formula.

When we evaluate (40) at (z,w), then the fact that (i_1^*, i_2^*) is an admissible pair implies that every term in the sum vanishes except for the term with $(u,u') = (i_1^*, i_2^*)$. So (40) evaluated at (z,w) becomes

$$\left(\partial_{i_1^*}\partial'_{i_2^*}(fg)\right)(z,w) = \left(\partial_{i_1^*}\partial'_{i_2^*} f\right)(z,w)\cdot g(z,w),$$

which is exactly the desired result. □

We are now ready to prove the main estimate of this section.

Proposition E.8.2. *Let s be a global section of $\mathcal{O}(\Omega(d_1, d_2, d))$ given by a collection of polynomials $\mathcal{F} = \{F_i\}$, and let (i_1^*, i_2^*) be an admissible pair for s. Then*

$$\begin{aligned} h_{C\times C,\Omega}(z,w) \geq -h(\mathcal{F}) &- (i_1^* + i_2^* + 2\delta_1 + 2\delta_2 + 2n) \\ &- \sum_v \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \sum_{k=1}^{\delta_1} \max_{0\leq \ell\leq n} \min_j \log \left| \left(\partial_{i_k} \frac{x_\ell}{x_j}\right)(z) \right|_v \\ &- \sum_v \max_{i'_1+\cdots+i'_{\delta_2}=i_2^*} \sum_{k=1}^{\delta_2} \max_{0\leq \ell\leq n} \min_{j'} \log \left| \left(\partial_{i'_k} \frac{x'_\ell}{x'_{j'}}\right)(w) \right|_v . \end{aligned}$$

PROOF. Recalling how we normalized the Weil height h_Ω in Section E.2 (see (8)), we have

$$\begin{aligned} h_{C\times C,\Omega}(z,w) &= \delta_1 h(\phi_{NA}(z)) + \delta_2 h(\phi_{NA}(w)) - dh(\phi_B(z,w)) \\ &= \delta_1 \sum_v \max_j \log\left|x_j(z)\right|_v + \delta_2 \sum_v \max_{j'} \log\left|x'_{j'}(w)\right|_v \\ &\quad - d\sum_v \max_i \log\left|y_i(z,w)\right|_v \\ &= -\Bigl(\delta_1 \sum_v \min_j \log\left|x_j(z)\right|_v^{-1} + \delta_2 \sum_v \min_{j'} \log\left|x'_{j'}(w)\right|_v^{-1} \\ &\quad + d\sum_v \max_i \log\left|y_i(z,w)\right|_v\Bigr) \\ &= -\sum_v \max_i \min_{j,j'} \log \left| \left(\frac{y_i^d}{x_j^{\delta_1} x'^{\,\delta_2}_{j'}} \right)(z,w) \right|_v . \end{aligned} \tag{41}$$

Note that (41) holds for any choice of projective coordinates

$$\phi_{NA}(z) = [x_0(z), \ldots, x_n(z)], \qquad \phi_{NA}(w) = [x'_0(w), \ldots, x'_n(w)],$$
$$\phi_B(z,w) = [y_0(z,w), \ldots, y_m(z,w)],$$

since the product formula will cancel out the effect of multiplying the coordinates by a nonzero scalar.

Next we want to use the fact that

$$\partial_{i_1^*}\partial'_{i_2^*} f(z,w) \neq 0.$$

We remind the reader that the standard way to exploit the fact that an algebraic number is not zero is to use the product formula, which after taking logarithms becomes

$$\sum_v \log \left|\partial_{i_1^*}\partial'_{i_2^*} f(z,w)\right|_v = 0. \tag{42}$$

Subtracting (42) from our lower bound (41) gives

$$h_{C\times C,\Omega}(z,w) \geq -\sum_v \max_i \min_{j,j'} \log \left| \left(\frac{y_i^d \partial_{i_1^*} \partial'_{i_2^*} f}{x_j^{\delta_1} x_{j'}^{\delta_2}} \right)(z,w) \right|_v . \tag{43}$$

It is more convenient to work with actual functions on $C \times C$, rather than homogeneous coordinates. Since we have assumed (39) that z and w do not lie on the coordinate hyperplanes, the product formula gives

$$\sum_v \log \left| \left(\frac{y_0^d}{x_0^{\delta_1} x_0'^{\delta_2}} \right)(z,w) \right|_v = 0, \tag{44}$$

where $x_0(z)$, $x_0'(w)$, and $y_0(z,w)$ are the same homogeneous coordinates being used in (43). Adding (44) to (43) gives

$$h_{C\times C,\Omega}(z,w) \geq -\sum_v \max_i \min_{j,j'} \log \left| \left(\frac{(y_i/y_0)^d \partial_{i_1^*} \partial'_{i_2^*} f}{(x_j/x_0)^{\delta_1} (x'_{j'}/x'_0)^{\delta_2}} \right)(z,w) \right|_v . \tag{45}$$

We are going to use Lemma E.8.1 to rewrite the $(v,i,j,j')^{\text{th}}$ term of this sum. If we take

$$g = \frac{(y_i/y_0)^d}{(x_j/x_0)^{\delta_1} (x'_{j'}/x'_0)^{\delta_2}}$$

in Lemma E.8.1 and use the definition of f, we obtain

$$\begin{aligned} \left(\frac{(y_i/y_0)^d \partial_{i_1^*} \partial'_{i_2^*} f}{(x_j/x_0)^{\delta_1} (x'_{j'}/x'_0)^{\delta_2}} \right)(z,w) &= \partial_{i_1^*} \partial'_{i_2^*} \left(\frac{(y_i/y_0)^d f}{(x_j/x_0)^{\delta_1} (x'_{j'}/x'_0)^{\delta_2}} \right)(z,w) \\ &= \left(\partial_{i_1^*} \partial'_{i_2^*} F_i \left(\frac{x}{x_j}, \frac{x'}{x'_{j'}} \right) \right)(z,w). \end{aligned}$$

Substituting this into (45) gives the comparatively neat lower bound

$$h_{C\times C,\Omega}(z,w) \geq -\sum_v \max_i \min_{j,j'} \log \left| \left(\partial_{i_1^*} \partial'_{i_2^*} F_i \left(\frac{x}{x_j}, \frac{x'}{x'_{j'}} \right) \right)(z,w) \right|_v , \tag{46}$$

and it remains only to estimate the size of the partial derivatives.

In order to keep the notation at a manageable level, we will first prove the following lemma.

Lemma E.8.3. *Let $\xi_0, \ldots, \xi_n \in K(C)$ be rational functions that are regular at z, and similarly let $\xi'_0, \ldots, \xi'_n \in K(C)$ be rational functions that are regular at w. Let*

$$F(\xi, \xi') = F(\xi_0, \ldots, \xi_n, \xi'_0, \ldots, \xi'_n)$$

be a bihomogeneous polynomial of bidegree (δ_1, δ_2). Let ζ and ζ' be uniformizers at z and w, respectively, and let ∂_i and ∂'_i be the differential operators (38) defined above. Then for any absolute value v,

$$\left|\left(\partial_{i_1^*}\partial'_{i_2^*}F(\xi,\xi')\right)(z,w)\right|_v \le 2_v^{i_1^*+i_2^*+2\delta_1+2\delta_2+2n}|F|_v \times \max_{i_1+\cdots+i_{\delta_1}=i_1^*}\prod_{k=1}^{\delta_1}\max_{0\le\ell\le n}\left|(\partial_{i_k}\xi_\ell)(z)\right|_v \times \max_{i'_1+\cdots+i'_{\delta_2}=i_2^*}\prod_{k=1}^{\delta_2}\max_{0\le\ell\le n}\left|(\partial_{i'_k}\xi'_\ell)(w)\right|_v.$$

Here $2_v = 2$ if v is archimedean, and $2_v = 1$ otherwise; and $|F|_v$ is the maximum of the absolute values of the coefficients of F.

PROOF. The number of monomials appearing in F is at most

$$\binom{\delta_1+n}{n}\binom{\delta_2+n}{n} \le 2^{\delta_1+\delta_2+2n}.$$

So using the triangle inequality gives

$$\left|\left(\partial_{i_1^*}\partial'_{i_2^*}F(\xi,\xi')\right)(z,w)\right|_v \le 2_v^{\delta_1+\delta_2+2n}\cdot|F|_v \times \max_{e_0+\cdots+e_n=\delta_1}\left|\left(\partial_{i_1^*}(\xi_0^{e_0}\cdots\xi_n^{e_n})\right)(z)\right|_v \times \max_{e'_0+\cdots+e'_n=\delta_2}\left|\left(\partial'_{i_2^*}(\xi_0'^{e'_0}\cdots\xi_n'^{e'_n})\right)(w)\right|_v. \tag{47}$$

In order to estimate the last two expressions on the right-hand side of (47), we look more generally at

$$\left|\left(\partial_I(\theta_1\cdots\theta_\delta)\right)(z)\right|_v,$$

for functions $\theta_1,\ldots,\theta_\delta \in K(C)$ that are regular at z. Leibniz's rule gives

$$\partial_I(\theta_1\cdots\theta_\delta) = \sum_{i_1+\cdots+i_\delta=I}(\partial_{i_1}\theta_1)\cdots(\partial_{i_\delta}\theta_\delta), \tag{48}$$

where again we note that the definition (38) of ∂ means that Leibniz's rule holds without the combinatorial symbols. The sum contains at most

$$\binom{I+\delta-1}{\delta-1} \le 2^{I+\delta}$$

terms, so evaluating (48) at z and using the triangle inequality gives

$$\left|\left(\partial_I(\theta_1\cdots\theta_\delta)\right)(z)\right|_v \le 2_v^{I+\delta}\max_{i_1+\cdots+i_\delta=I}\prod_{k=1}^{\delta}\left|(\partial_{i_k}\theta_k)(z)\right|_v. \tag{49}$$

Now apply (49) to the expressions on the right-hand side of (47). For example, taking $I = i_1^*$, $\delta = \delta_1$, and $\theta_1 \cdots \theta_\delta = \xi_0^{e_0} \cdots \xi_n^{e_n}$ gives

$$\max_{e_0+\cdots+e_n=\delta_1} \left|\left(\partial_{i_1^*}(\xi_0^{e_0}\cdots\xi_n^{e_n})\right)(z)\right|_v \le 2_v^{i_1^*+\delta_1} \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \prod_{k=1}^{\delta_1} \left|(\partial_{i_k}\theta_k)(z)\right|_v,$$

where each θ_k equals one of $\xi_0,\ldots,\xi_n$. Thus

$$\begin{aligned}\max_{e_0+\cdots+e_n=\delta_1} &\left|\left(\partial_{i_1^*}(\xi_0^{e_0}\cdots\xi_n^{e_n})\right)(z)\right|_v \\ &\le 2_v^{i_1^*+\delta_1} \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \prod_{k=1}^{\delta_1} \max_{0\le \ell\le n} \left|(\partial_{i_k}\xi_\ell)(z)\right|_v.\end{aligned} \tag{50}$$

In a similar fashion we derive the upper bound

$$\begin{aligned}\max_{e'_0+\cdots+e'_n=\delta_2} &\left|\left(\partial'_{i_2^*}({\xi'_0}^{e'_0}\cdots{\xi'_n}^{e'_n})\right)(w)\right|_v \\ &\le 2_v^{i_2^*+\delta_2} \max_{i'_1+\cdots+i'_{\delta_2}=i_2^*} \prod_{k=1}^{\delta_2} \max_{0\le \ell\le n} \left|(\partial_{i'_k}\xi'_\ell)(w)\right|_v.\end{aligned} \tag{50$'$}$$

Substituting (50) and (50′) into (47) and doing a little algebra completes the proof of Lemma E.8.3. □

We now resume the proof of Proposition E.8.2. Recall that we have derived the lower bound (46), which we repeat here for the convenience of the reader:

$$h_{C\times C,\Omega}(z,w) \ge -\sum_v \max_i \min_{j,j'} \log \left|\left(\partial_{i_1^*}\partial'_{i_2^*}F_i\left(\frac{x}{x_j}, \frac{x'}{x'_{j'}}\right)\right)(z,w)\right|_v. \tag{46}$$

We apply Lemma E.8.3 with $F = F_i$, $\xi_\ell = x_\ell/x_j$, and $\xi'_\ell = x'_\ell/x'_{j'}$, which yields

$$\begin{aligned}\left|\left(\partial_{i_1^*}\partial'_{i_2^*}F_i\left(\frac{x}{x_j}, \frac{x'}{x'_{j'}}\right)\right)(z,w)\right|_v &\le 2_v^{i_1^*+i_2^*+2\delta_1+2\delta_2+2n}|F_i|_v \\ &\quad\times \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \prod_{k=1}^{\delta_1} \max_{0\le\ell\le n} \left|\left(\partial_{i_k}\frac{x_\ell}{x_j}\right)(z)\right|_v \\ &\quad\times \max_{i'_1+\cdots+i'_{\delta_2}=i_2^*} \prod_{k=1}^{\delta_2} \max_{0\le\ell\le n} \left|\left(\partial_{i'_k}\frac{x'_\ell}{x'_{j'}}\right)(w)\right|_v.\end{aligned}$$

Finally, we substitute this into (46) to obtain

$$\begin{aligned} h_{C\times C,\Omega}(z,w) \geq -(i_1^*+i_2^*+2\delta_1+2\delta_2+2n)\log 2 - h(\mathcal{F}) \\ -\sum_v \min_j \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \sum_{k=1}^{\delta_1} \max_{0\leq \ell\leq n} \log\left|\left(\partial_{i_k}\frac{x_\ell}{x_j}\right)(z)\right|_v \\ -\sum_v \min_{j'} \max_{i'_1+\cdots+i'_{\delta_2}=i_2^*} \sum_{k=1}^{\delta_2} \max_{0\leq \ell\leq n} \log\left|\left(\partial_{i'_k}\frac{x'_\ell}{x'_{j'}}\right)(w)\right|_v, \end{aligned}$$

which is slightly stronger than the desired result. This completes the proof of Proposition E.8.2. □

E.9. Eisenstein's Estimate for the Derivatives of an Algebraic Function

In this section we will derive an upper bound for the quantities

$$\left|\left(\partial_{i_k}\frac{x_\ell}{x_j}\right)(z)\right|_v \qquad \text{and} \qquad \left|\left(\partial_{i'_k}\frac{x'_\ell}{x'_{j'}}\right)(w)\right|_v$$

appearing in Proposition E.8.2. Recall that the function x_ℓ/x_j is a rational function on C via the embedding $\phi_{NA}: C \hookrightarrow \mathbb{P}^n$, and that ∂_i involves differentiating with respect to a uniformizer ζ at $z \in C$. An alternative way to treat this situation is to consider the finite map $\zeta: C \longrightarrow \mathbb{P}^1$. This means that $K(C)$ is a finite extension of $K(\zeta)$, so any rational function ξ on C satisfies a polynomial equation $p(\xi,\zeta) = 0$ of degree at most the degree of $K(C)$ over $K(\zeta)$. If further the partial derivative p_ξ does not vanish at z, then the implicit function theorem says that we can write ξ as an analytic function of ζ in a neighborhood of z. Thus

$$\xi(\zeta) = \sum_{i\geq 0} (\partial_i \xi(z))\zeta^i, \tag{51}$$

where as usual we are writing $\partial_i = \dfrac{1}{i!}\left(\dfrac{\partial}{\partial\zeta}\right)^i$.

Suppose now that we are working over the complex numbers. A basic result from complex analysis says that (51) converges for all $|\zeta| < \rho$, where

$$\rho = \liminf_{i\to\infty} \sqrt[i]{|\partial_i \xi(z)|^{-1}}.$$

Further, the radius of convergence ρ satisfies $\rho > 0$. It follows that if we replace ρ by a smaller $\rho_1 > 0$, we get

$$\left|\partial_i \xi(z)\right| \le c\rho_1^{-i} \qquad \text{for all } i \ge 0, \tag{52}$$

where the constant c is thrown in to take care of the first few i's.

This is exactly the sort of estimate we need for Proposition E.8.2, except that as usual we must keep track of the dependence on the various quantities involved. The next proposition gives a proof of (52) with ρ_1 given explicitly in terms of the polynomial $p(\xi, \zeta)$.

Proposition E.9.1. *Let $p(\xi,\zeta) \in K[\xi,\zeta]$ be a polynomial of degree D, and let $a \in K$ be a value such that $p(a,0) = 0$ and such that the partial derivative $p_\xi(a,0)$ is nonzero. Let $\xi = \xi(z)$ be the algebraic function satisfying $p(\xi(\zeta),\zeta) = 0$ and $\xi(0) = a$. Then for each absolute value on K, the Taylor coefficients of this algebraic function satisfy*

$$\left|\partial_i \xi(0)\right|_v \le (2D)_v^{11i} \left(\frac{|p|_v}{\left|p_\xi(a,0)\right|_v}\right)^{2i-1} \max\left\{1, |a|_v\right\}^{2iD}.$$

Here we are writing

$$\begin{aligned} p_\xi &= \partial p/\partial \xi, \\ |p|_v &= \max |\textit{coefficients of } p|_v, \\ (2D)_v &= \begin{cases} |2D|_v & \textit{if } v \textit{ is archimedean,} \\ 1 & \textit{otherwise.} \end{cases} \end{aligned}$$

PROOF. We claim that for each $i \ge 1$ there is a polynomial $q_i(\xi,\zeta)$ such that

$$q_i + (p_\xi)^{2i-1} \frac{\partial^i \xi}{\partial \zeta^i} = 0. \tag{53}$$

To see this, we begin by differentiating $p(\xi,\zeta) = 0$ to obtain

$$p_\zeta + p_\xi \frac{\partial \xi}{\partial \zeta} = 0. \tag{54}$$

This gives (53) for $i = 1$ with $q_1 = p_\zeta$. Next suppose that we know (53) for i. Then differentiating (53) yields

$$(q_i)_\xi \xi_\zeta + (q_i)_\zeta + (2i-1)(p_\xi)^{2i-2}\left(p_{\xi\xi}\xi_\zeta + p_{\xi\zeta}\right)\frac{\partial^i \xi}{\partial \zeta^i} + (p_\xi)^{2i-1}\frac{\partial^{i+1}\xi}{\partial \zeta^{i+1}} = 0. \tag{55}$$

Now we use (53) (which is true for i by hypothesis) and (54) to eliminate $\partial^i \xi/\partial \zeta^i$ and ξ_ζ from (55). Note that this is all right in a neighborhood

of $\zeta = 0$, since we have assumed that $p_\xi(a,0) \neq 0$. Clearing denominators, we find that

$$-(q_i)_\xi p_\zeta p_\xi + (q_i)_\zeta p_\xi^2 + (2i-1)q_i\left(-p_{\xi\xi}p_\zeta + p_{\xi\zeta}p_\xi\right) + (p_\xi)^{2i+1}\frac{\partial^{i+1}\xi}{\partial\zeta^{i+1}} = 0.$$

This proves by induction that (53) is valid for all $i \geq 1$, with the additional information that the q_i's are given by the recursive formula

$$\begin{aligned} q_1 &= p_\zeta, \\ q_{i+1} &= -(q_i)_\xi p_\zeta p_\xi + (q_i)_\zeta p_\xi^2 + (2i-1)q_i\left(-p_{\xi\xi}p_\zeta + p_{\xi\zeta}p_\xi\right). \end{aligned} \tag{56}$$

We can easily use (56) to estimate the degrees of the q_i's. Thus

$$\deg q_1 = \deg p_\zeta \leq \deg p - 1 = D - 1,$$

and

$$\begin{aligned} \deg q_{i+1} &\leq \max\left\{\deg\big((q_i)_\xi p_\zeta p_\xi\big), \deg\big((q_i)_\zeta p_\xi^2\big), \deg\big(q_i p_{\xi\xi}p_\zeta\big), \deg\big(q_i p_{\xi\zeta}p_\xi\big)\right\} \\ &\leq \deg q_i + 2D - 3. \end{aligned}$$

Applying this repeatedly gives

$$\deg q_i \leq (2i-1)D - (3i-2) \leq (2i-1)(D-1). \tag{57}$$

In a similar fashion we can estimate the height of the q_i's using the elementary height estimates proven in Section B.7. Thus applying (B.7.4(b)) to the four terms in the recursion (56) gives

$$\begin{aligned} &|q_{i+1}|_v \\ &\leq 4_v \max\left\{\left|(q_i)_\xi p_\zeta p_\xi\right|_v, \left|(q_i)_\zeta p_\xi^2\right|_v, \left|(2i-1)q_i p_{\xi\xi}p_\zeta\right|_v, \left|(2i-1)q_i p_{\xi\zeta}p_\xi\right|_v\right\}. \end{aligned} \tag{58}$$

(As usual, N_v is an abbreviation for $|N|_v$ if v is archimedean, and $N_v = 1$ if v is nonarchimedean.)

Now we estimate each of the four quantities on the right-hand side of (58). First,

$$\begin{aligned} \left|(q_i)_\xi p_\zeta p_\xi\right|_v &\leq (2\deg p_\zeta)_v^2 (2\deg p_\zeta)_v^2 \left|(q_i)_\xi\right|_v |p_\zeta|_v |p_\xi|_v && \text{from (B.7.4(a))} \\ &\leq (2D-2)_v^4 (\deg q_i)_v |q_i|_v D_v^2 |p|_v && \text{from (B.7.4(c))} \\ &\leq (2D-2)_v^4 (2i-1)_v (D-1)_v D_v^2 |q_i|_v |p|_v^2 && \text{from (57)} \\ &\leq 2_v^5 D_v^7 i_v |q_i|_v |p|_v^2. \end{aligned}$$

Similarly,

$$\left|(q_i)_\zeta p_\xi^2\right|_v \leq 2_v^5 D_v^7 i_v |q_i|_v |p|_v^2.$$

Next,

$$\begin{aligned}
&\bigl|(2i-1)q_i p_{\xi\xi} p_\zeta\bigr|_v \\
&\quad\le (2i-1)_v (2\deg p_{\xi\xi})_v^2 (2\deg p_\zeta)_v^2 |q_i|_v |p_{\xi\xi}|_v |p_\zeta|_v \quad \text{from (B.7.4(a))}\\
&\quad\le (2i-1)_v (2D-4)_v^2 (2D-2)_v^2 |q_i|_v D_v^3 |p|_v^2 \quad \text{from (B.7.4(c))}\\
&\quad\le 2_v^5 D_v^7 i_v |q_i|_v |p|_v^2.
\end{aligned}$$

Finally, a similar calculation gives the same estimate for the fourth quantity in (58),

$$\bigl|(2i-1)q_i p_{\xi\zeta} p_\xi\bigr|_v \le 2_v^5 D_v^7 i_v |q_i|_v |p|_v^2.$$

Using these last four inequalities, (58) becomes

$$|q_{i+1}|_v \le 2_v^7 D_v^7 i_v |p|_v^2 |q_i|_v. \tag{59}$$

Applying (59) repeatedly and using $|q_1|_v = |p_\xi|_v \le D_v |p|_v$ gives

$$\begin{aligned}
|q_i|_v &\le (2D)_v^{7(i-1)} \bigl((i-1)!\bigr)_v |p|_v^{2(i-1)} |q_1|_v \\
&\le (2D)_v^{7i} \bigl((i-1)!\bigr)_v |p|_v^{2i-1}.
\end{aligned} \tag{60}$$

We are finally ready to estimate the size of the partial derivatives of ξ. Evaluating (53) at $\zeta = 0$, we obtain

$$\bigl|\partial_i \xi(0)\bigr|_v = \left(\frac{1}{i!}\right)_v \left|\frac{\partial^i \xi}{\partial \zeta^i}(0)\right|_v = \left(\frac{1}{i!}\right)_v \left|-\frac{q_i(a,0)}{p_\xi(a,0)^{2i-1}}\right|_v. \tag{61}$$

(Note that $\xi(0) = a$.) We can bound the numerator by calculating

$$\begin{aligned}
&\bigl|q_i(a,0)\bigr|_v \le (2\deg q_i)_v^2 |q_i|_v \max\bigl\{1, |a|_v\bigr\}^{\deg q_i} \quad \text{from (B.7.4(d))}\\
&\le \bigl(2(2i-1)(D-1)\bigr)_v^2 (2D)_v^{7i} \bigl((i-1)!\bigr)_v |p|_v^{2i-1} \max\bigl\{1, |a|_v\bigr\}^{(2i-1)(D-1)} \\
&\qquad\qquad \text{from (57) and (60)}\\
&\le 2_v^{7i+4} D_v^{7i+2} i_v (i!)_v |p|_v^{2i-1} \max\bigl\{1, |a|_v\bigr\}^{2iD} \\
&\le (2D)_v^{11i} (i!)_v |p|_v^{2i-1} \max\bigl\{1, |a|_v\bigr\}^{2iD}.
\end{aligned} \tag{62}$$

For the last line we have used the fact that $i \ge 1$ and trivial estimates such as $i \le 2^{i-1}$. Substituting (62) into (61) gives

$$\bigl|\partial_i \xi(0)\bigr|_v \le (2D)_v^{11i} \left(\frac{|p|_v}{\bigl|p_\xi(a,0)\bigr|_v}\right)^{2i-1} \max\bigl\{1, |a|_v\bigr\}^{2iD}, \tag{63}$$

which completes the proof of Proposition E.9.1. □

Note that for almost all nonarchimedean places v we obtain the bound $\left|\partial_i\xi(0)\right|_v \le 1$; in other words, $\partial_i\xi(0)$ is v-integral. It is worth noting that when we use (B.7.4(a)) to estimate the four quantities in (58), it is essential that the estimate $\left|\prod f_j\right|_v \le c_v \prod |f_j|_v$ hold for a constant c_v that is independent of the first polynomial f_1. If this had not been true, then (59) would have looked like $|q_{i+1}|_v \ll i_v^3|p|_v^2|q_i|_v$, and so our upper bound (60) for $|q_i|_v$ would have involved $(i-1)!^3$. Since the $i!$ in the definition of ∂_i is able to cancel only one of these factorials, our final estimate (63) would have been multiplied by $(i!)_v^2$. This may not seem too terrible, since when we take the logarithm of (63), the difference between having $\log(2D)^{11i} \approx i \log D$ and $\log i!^2 \approx i \log i$ is only an additional factor of $\log i$. However, this would change the bound in Proposition E.10.1 (to be proven in the next section) from

$$c_{16}(i_1^*|z|^2 + i_2^*|w|^2) + c_{17}(i_1^* + i_2^*)$$

to

$$c_{16}(i_1^*|z|^2 + i_2^*|w|^2) + c_{17}(i_1^* \log i_1^* + i_2^* \log i_2^*),$$

and the extra factor of $\log i_1^*$ would eventually prevent us from proving Vojta's inequality. This gives some indication of how delicately the various estimates must fit together if the proof of Vojta's inequality is to succeed.

E.10. Lower Bound for $h_\Omega(z,w)$ at Admissible (i_1^*, i_2^*): Version II

In Section E.8 we proved a lower bound for $h_\Omega(z,w)$ in terms of an admissible pair of indices (i_1^*, i_2^*). This lower bound depends on the partial derivatives of certain algebraic functions on C. We are now going to use Proposition E.9.1 to estimate those partial derivatives and prove the following improved lower bound for $h_\Omega(z,w)$.

Proposition E.10.1. *Let s be a global section of $\mathcal{O}(\Omega(d_1,d_2,d))$ given by a collection of polynomials $\mathcal{F} = \{F_i\}$, and let (i_1^*, i_2^*) be an admissible pair for s. There is a finite set of points $\mathcal{Z} \subset C(\bar{K})$ such that for all points $z, w \in C(\bar{K})$ with $z, w \notin \mathcal{Z}$,*

$$h_{C\times C,\Omega}(z,w) \ge -h(\mathcal{F}) - c_{18}\big(i_1^*|z|^2 + i_2^*|w|^2\big) - c_{19}(i_1^* + i_2^* + \delta_1 + \delta_2 + 1).$$

PROOF. Recall (see Section E.7) that the projection map

$$\begin{array}{rccc} \pi : C & \xrightarrow{\phi_{NA}} & \mathbb{P}^n & \longrightarrow & \mathbb{P}^1, \\ & & x & \longmapsto & [x_0, x_1], \end{array}$$

is a finite morphism of degree N. It follows that for any indices ℓ, j, the rational function $x_\ell/x_j \in K(C)$ is algebraic over $K\big(x_1/x_0\big)$, and we would like to estimate the degree of the corresponding algebraic relation.

To do this, we look at the composition of maps

$$\begin{array}{cccccccccc} C & \xrightarrow{\phi_{NA}} & \mathbb{P}^n & \xrightarrow{\psi_{\ell j}} & \mathbb{P}^1\times\mathbb{P}^1 & \xrightarrow{\text{Segre}} & \mathbb{P}^3 & \xrightarrow{\text{proj.}} & \mathbb{P}^2, \\ & & x & \longmapsto & \big([x_\ell,x_j],[x_0,x_1]\big), & & & & \\ & & & & \big([a,b],[c,d]\big) & \longmapsto & [ac,bd,bc,ad], & & \\ & & & & & & [x,y,z,w] & \longmapsto & [x,y,z]. \end{array}$$

Now start with a line in $\mathbb{P}^2$ and pull it back step by step to C. Assuming that $\ell \neq j$ and that $(\ell,j) \neq (1,0),\ (0,1)$, the line in $\mathbb{P}^2$ pulls back to a hyperplane in $\mathbb{P}^3$, then to a pair of transversal lines in $\mathbb{P}^1\times\mathbb{P}^1$. These, in turn, pull back to two hyperplanes in $\mathbb{P}^n$, which finally gives a divisor of degree $2N$ on C. In conclusion, the map

$$C \xrightarrow{[x_\ell x_0,\,x_j x_1,\,x_j x_0]} \mathbb{P}^2$$

embeds C as a curve of degree $2N$ in $\mathbb{P}^2$, so there is a homogeneous polynomial $G_{\ell j}$ of degree $2N$ such that

$$G_{\ell j}(x_\ell x_0, x_j x_1, x_j x_0) = 0 \qquad \text{on } C.$$

Dividing this by $(x_j x_0)^{2N}$ gives a polynomial relation

$$g_{\ell j}\left(\frac{x_\ell}{x_j},\frac{x_1}{x_0}\right) = 0 \qquad \text{on } C,$$

where $g_{\ell j}$ is a polynomial of degree at most $2N$ with coefficients in K.

Taking a point $z \in C(\bar{K})$ as usual, we define shifted polynomials

$$p_{\ell j}(S,T) = g_{\ell j}\left(S, T + \frac{x_1}{x_0}(z)\right).$$

From (B.7.4(e)) we have

$$|p_{\ell j}|_v = \left|g_{\ell j}\left(S,T+\frac{x_1}{x_0}(z)\right)\right|_v \le 2_v^{2\deg g_{\ell j}} |g_{\ell j}|_v \max\left\{1, \left|\frac{x_1}{x_0}(z)\right|_v\right\}^{\deg g_{\ell j}}.$$

Since there are only finitely many choices for ℓ, j, and since each $g_{\ell j}$ has degree at most $2N$, we obtain

$$|p_{\ell j}|_v \le c_{20}(v) \max\left\{1, \left|\frac{x_1}{x_0}(z)\right|_v\right\}^{2N}, \tag{64}$$

where the constant $c_{20}(v)$ satisfies $c_{20}(v) \geq 1$ for all v, and $c_{20}(v) = 1$ for all but finitely many v. Constants of this sort are called (multiplicative) M_K-constants; see Section B.8 or Lang [6].

Now we apply Proposition 12.1 with

$$i = i_k, \qquad \xi = \frac{x_\ell}{x_j}, \qquad \zeta = \frac{x_1}{x_0} - \frac{x_1}{x_0}(z),$$
$$p = p_{\ell j}, \qquad a = \xi(0) = \frac{x_\ell}{x_j}(z), \qquad D = \deg p_{\ell j} \leq 2N.$$

Note that for all but finitely many points of $z \in C$, the function ζ will be a uniformizer at z. Similarly, since the curve $g_{\ell j} = 0$ in $\mathbb{P}^2$ is a (possibly singular) model for C, there are only finitely many points of C at which its partial derivative $(g_{\ell j})_\xi$ vanishes, where we are writing $(g_{\ell j})_\xi$ for the partial derivative with respect to the first variable. So in the statement of Proposition E.10.1, we take $\mathcal{Z}$ to be the finite set of points

$$\mathcal{Z} = \left\{ \begin{aligned} u \in C(\bar{K}) \ : \ &\frac{x_1}{x_0} - \frac{x_1}{x_0}(u) \text{ is not a uniformizer at } u \text{, or} \\ &(g_{\ell j})_\xi \left(\frac{x_\ell}{x_j}(u), \frac{x_1}{x_0}(u) \right) = 0 \text{ for some } \ell, j \end{aligned} \right\},$$

and we assume throughout that $z, w \notin \mathcal{Z}$. Using Proposition E.9.1, we obtain the estimate

$$\left| \left(\partial_{i_k} \frac{x_\ell}{x_j} \right)(z) \right|_v \leq (4N)_v^{11 i_k} \left(\frac{|p_{\ell j}|_v}{\left|(p_{\ell j})_\xi(a,0)\right|_v} \right)^{2i_k - 1} \max\left\{ 1, \left| \frac{x_\ell}{x_j}(z) \right|_v \right\}^{4 i_k N}. \tag{65}$$

Next we take the minimum over $0 \leq j \leq n$. Note that for any given point z, there will be at least one index j such that

$$\left| \frac{x_\ell}{x_j}(z) \right|_v \leq 1 \qquad \text{for all } 0 \leq \ell \leq n.$$

So taking the minimum of (65) and using $\min\{a_j b_j\} \leq \max\{a_j\} \min\{b_j\}$ gives

$$\min_{0 \leq j \leq n} \left| \left(\partial_{i_k} \frac{x_\ell}{x_j} \right)(z) \right|_v \leq (4N)_v^{11 i_k} \max_{0 \leq j \leq n} \left(\frac{|p_{\ell j}|_v}{\left|(p_{\ell j})_\xi(a,0)\right|_v} \right)^{2i_k - 1}. \tag{66}$$

Next we observe that there are only finitely many polynomials $p_{\ell j}$, so

$$|p_{\ell j}|_v \leq c_{21}(v). \tag{67}$$

Further, referring to the definition of $p_{\ell j}$ in terms of $g_{\ell j}$, we see that

$$(p_{\ell j})_\xi(a, 0) = (g_{\ell j})_\xi \left(\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z) \right), \tag{68}$$

where as above we are writing $(g_{\ell j})_\xi$ for the partial derivative of $g_{\ell j}$ with respect to its first variable. Substituting (67) and (68) into (66) gives

$$\min_{0\le j\le n}\left|\left(\partial_{i_k}\frac{x_\ell}{x_j}\right)(z)\right|_v \le (4N)_v^{11i_k}\max_{0\le j\le n}\left(\frac{c_{21}(v)}{\left|(g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z),\frac{x_1}{x_0}(z)\right)\right|_v}\right)^{2i_k-1}. \tag{69}$$

It is time to look back at Proposition E.8.2, which gives the lower bound

$$h_{C\times C,\Omega}(z,w)\ge -h(\mathcal{F})-(i_1^*+i_2^*+2\delta_1+2\delta_2+2n)-M(z,i_1^*)-M'(w,i_2^*). \tag{70}$$

Here $M(z,i_1^*)$ is the "messy" expression

$$M(z,i_1^*)=\sum_v \max_{i_1+\cdots+i_{\delta_1}=i_1^*}\sum_{k=1}^{\delta_1}\max_{0\le \ell\le n}\min_j \log\left|\left(\partial_{i_k}\frac{x_\ell}{x_j}\right)(z)\right|_v, \tag{71}$$

and $M'(w,i_2^*)$ is defined similarly. Since $M(z,i_1^*)$ and $M'(w,i_2^*)$ appear with negative signs, we need to find upper bounds. Using (69), we see that

$$M(z,i_1^*)\le \sum_v \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \sum_{k=1}^{\delta_1}\max_{0\le \ell,j\le n}\log\left[(4N)_v^{11i_k}\left(\frac{c_{21}(v)}{\left|(g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z),\frac{x_1}{x_0}(z)\right)\right|_v}\right)^{2i_k-1}\right]. \tag{72}$$

We estimate the inner logarithm in (72) using a sum of three terms

$$\begin{aligned}\max_{0\le \ell,j\le n}\log&\left[(4N)_v^{11i_k}\left(\frac{c_{21}(v)}{\left|(g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z),\frac{x_1}{x_0}(z)\right)\right|_v}\right)^{2i_k-1}\right]\\ &\le 11i_k\log(4N)_v+(2i_k-1)\log c_{21}(v)\\ &\quad +(2i_k-1)\sum_{0\le \ell,j\le n}\log^+\left|(g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z),\frac{x_1}{x_0}(z)\right)\right|_v^{-1}\end{aligned} \tag{73}$$

and bound each term separately. Here we are using the standard notation

$$\log^+ t=\max\{0,\log t\}\qquad \text{for } t\in\mathbb{R},\ t\ge 0.$$

The first two terms in (73) are easy to estimate. Using the fact that $(4N)_v=1$ for nonarchimedean v and $c_{21}(v)=1$ for all but finitely many v,

we obtain

$$\sum_v \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \sum_{k=1}^{\delta_1} 11 i_k \log(4N)_v = 11 i_1^* \log 4N \le c_{22} i_1^*, \tag{74}$$

$$\sum_v \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \sum_{k=1}^{\delta_1} (2i_k - 1)\log c_{21}(v) = (2i_1^* - 1)\sum_v \log c_{21}(v) \le c_{23} i_1^*. \tag{75}$$

It remains to bound the third term in (73). We begin by computing

$$\begin{aligned}
&\sum_v \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \sum_{k=1}^{\delta_1} (2i_k - 1) \sum_{0\le \ell,j\le n} \log^+ \left|(g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z)\right)\right|_v^{-1} \\
&= (2i_1^* - 1) \sum_{0\le \ell,j\le n} \sum_v \log^+ \left|(g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z)\right)\right|_v^{-1} \\
&= (2i_1^* - 1) \sum_{0\le \ell,j\le n} h\left((g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z)\right)\right) \\
&\qquad\qquad \text{from the definition of height and the fact that } h(\alpha^{-1}) = h(\alpha) \text{ for } \alpha \ne 0 \\
&\le (2i_1^* - 1) \sum_{0\le \ell,j\le n} \left\{ \deg\big((g_{\ell j})_\xi\big) h\left(\left[\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z), 1\right]\right)\right. \qquad (76) \\
&\qquad\qquad \left. + h\big((g_{\ell j})_\xi\big) + 3\log\big(3 + \deg\big((g_{\ell j})_\xi\big)\big) \right\} \\
&\qquad\qquad \text{from (B.7.1(b)).}
\end{aligned}$$

Since there are only finitely many $g_{\ell j}$'s, and since they do not depend on the point z or the index i_1^*, we have, using the elementary degree and height estimates for the derivative of a polynomial (Proposition B.7.4),

$$\deg(g_{\ell j})_\xi \le 2N - 1 \le c_{24} \qquad \text{and} \qquad h\big((g_{\ell j})_\xi\big) \le h(g_{\ell j}) + \log\deg g_{\ell j} \le c_{25}. \tag{77}$$

To estimate the other term in (76), we use the fact that for any point $[\alpha_0, \dots, \alpha_n] \in \mathbb{P}^n$,

$$h\big([\dots, \alpha_i\alpha_j, \dots]_{0\le i\le j\le n}\big) = 2h\big([\alpha_0, \dots, \alpha_n]\big).$$

(This is the 2-uple embedding $\mathbb{P}^n \hookrightarrow \mathbb{P}^{(n^2+3n)/2}$, and we apply the height estimate for d-uple embedding (B.2.4c).) So for any indices ℓ, j such that $\alpha_0\alpha_j \ne 0$, we have

$$h\left(\left[\frac{\alpha_l}{\alpha_j}, \frac{\alpha_1}{\alpha_0}, 1\right]\right) = h\big([\alpha_0\alpha_\ell, \alpha_1\alpha_j, \alpha_0\alpha_j]\big) \le 2h\big([\alpha_0, \dots, \alpha_n]\big).$$

Applying this to $\phi_{NA}(z) = [x_0(z),\dots,x_n(z)]$ gives

$$h\left(\left[\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z), 1\right]\right) \le 2h\left([x_0(z),\dots,x_n(z)]\right) = 2h(\phi_{NA}(z)). \tag{78}$$

We need to relate this last quantity to the height $|z|^2$. We already did this in Section E.4, but we will briefly recall the details. The height $|z|^2$ is defined by $|z|^2 = \hat{h}_{J,\Theta_A}(j_A(z))$, where Θ_A is the Θ-divisor on J and $j_A : C \to J$ is the embedding described in Section E.2. Lemma E.2.1(b) says that we have the linear equivalence $j_A^*\Theta_A \sim gA$, so the Height Machine says that for all $u \in C(\bar{K})$,

$$\begin{aligned} |u|^2 &= \hat{h}_{J,\Theta_A}(j_A(u)) \\ &= h_{J,\Theta_A}(j_A(u)) + O(1) \\ &= h_{C,j_A^*\Theta_A}(u) + O(1) \\ &= gh_{C,A}(u) + O(1) \\ &= \frac{g}{N}h_{C,NA}(u) + O(1) \\ &= \frac{g}{N}h(\phi_{NA}(u)) + O(1). \end{aligned}$$

Taking $u = z$ and substituting into (78) gives

$$h\left(\left[\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z), 1\right]\right) \le \frac{2N}{g}|z|^2 + c_{26}. \tag{79}$$

Now we substitute (77) and (79) into (76), which gives the estimate

$$\begin{aligned} \sum_v \max_{i_1+\cdots+i_{\delta_1}=i_1^*} \sum_{k=1}^{\delta_1}(2i_k-1) \sum_{0\le\ell,j\le n} \log^+ \left|(g_{\ell j})_\xi\left(\frac{x_\ell}{x_j}(z), \frac{x_1}{x_0}(z)\right)\right|_v^{-1} \\ \le (2i_1^*-1)\sum_{0\le\ell,j\le n}\left(c_{27}|z|^2 + c_{28}\right) \\ \le c_{29}i_1^*|z|^2 + c_{30}i_1^*. \end{aligned} \tag{80}$$

We have now estimated each of the three inner logarithmic terms in (72). More precisely, we substitute (73) into (72), break the result into three sums, and estimate those three sums by using (74), (75), and (80). This gives

$$M(z,i_1^*) \le c_{31}i_1^*|z|^2 + c_{32}i_1^*. \tag{81}$$

A similar calculation gives an analogous estimate for the other "messy" expression $M'(w,i_2^*)$ appearing in the statement of Proposition E.8.2,

$$M'(w,i_2^*) \le c_{33}i_2^*|w|^2 + c_{34}i_2^*. \tag{82}$$

Substituting (81) and (82) into (70) and combining some of the terms gives the desired result, which completes the proof of Proposition E.10.1. □

Remark E.10.2. A careful analysis of the proof of Proposition E.10.1 shows that the constant c_{18} may be chosen to depend only on g. More precisely, the computations (with slightly different notation) of de Diego [1, Lemme 5.4] say that the value $c_{18} = 12n^2N^2/g$ is admissible.

E.11. A Nonvanishing Derivative of Small Order

Our aim in this section is to show that the small section s constructed in Section E.7 has a small derivative that does not vanish at (z, w). In other words, we will prove that s has a small admissible pair (i_1^*, i_2^*). We begin with a nonvanishing result for ordinary polynomials.

Proposition E.11.1. (Two-variable Roth's lemma) *Let $P \in \bar{\mathbb{Q}}[X_1, X_2]$ be a nonzero polynomial of degree at most r_1 in X_1 and at most r_2 in X_2, and let $\beta_1, \beta_2 \in \bar{\mathbb{Q}}$. Suppose that $1 \geq \omega > 0$ is a constant such that*

$$r_2 \leq \omega r_1 \qquad \textit{and} \qquad h(P) + 4r_1 \leq \omega \min\{r_1 h(\beta_1), r_2 h(\beta_2)\}.$$

Then there are indices $i_1, i_2 \geq 0$ such that

$$\frac{i_1}{r_1} + \frac{i_2}{r_2} \leq 4\sqrt{\omega} \qquad \text{and} \qquad \frac{\partial^{i_1+i_2} P}{\partial^{i_1} X_1 \partial^{i_2} X_2}(\beta_1, \beta_2) \neq 0. \tag{83}$$

Remark. It is actually a misnomer to call Proposition E.11.1 "Roth's lemma," since versions were known to Siegel, Gelfand, and Dyson. Roth's contribution was to prove the appropriate generalization of (E.11.1) to the case of polynomials in more than two variables. But just as "Siegel's lemma" is now often used to refer to any result in which one estimates the size of solutions to a system of linear equations, we will use "Roth's lemma" as a generic description of any nonvanishing result of the sort described in Proposition E.11.1.

PROOF (of Proposition E.11.1). This is simply Roth's lemma (D.6.2) with $m = 2$ (i.e., two variables) and $\omega = \eta^2$. This completes the proof of Proposition E.11.1. □

We now apply Roth's lemma to show that a small global section associated to a Vojta divisor admits a small admissible pair.

Proposition E.11.2. *There is a constant $c_{35} = c_{35}(C, \phi_{NA}, \phi_B)$ such that the following is true. Let $0 < \varepsilon, \gamma < 1$ be small constants, and suppose that the integers d_1, d_2, d and the points $z, w \in C(\bar{K})$ satisfy the following conditions:*

$$\varepsilon^2 d_1 \geq d_2 \qquad \text{and} \qquad \min\{d_2|w|^2, d_1|z|^2\} \geq \frac{c_{35}}{\gamma\varepsilon^2} d_1. \tag{84}$$

Suppose further that $d_1d_2 - gd^2 \geq \gamma d_1 d_2$, and let s be a small global section of $\mathcal{O}\big(\Omega(d_1, d_2, d)\big)$ as described in Proposition E.7.1. Then there exists an admissible pair (i_1^*, i_2^*) for s with

$$\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2} \leq 12N\varepsilon.$$

PROOF. We proved earlier (see equation (10) in Section E.4) that for any point $u \in C(\bar{K})$, we have

$$h_{C,A}(u) = \frac{1}{g}|u|^2 + O(1). \tag{85}$$

Next we recall the affine coordinate functions $\xi_i = (x_i/x_0)\big|_C$ defined in Section E.7. We will write (ξ, ξ') for the analogous affine coordinates on $C \times C$. Then for any point $u \in C$ with $x_0(u) \neq 0$, we have

$$h_{C,NA}(u) = h\big(\phi_{NA}(u)\big) = h\big([1, \xi_1(u), \ldots, \xi_n(u)]\big) \leq \sum_{i=1}^{n} h\big(\xi_i(u)\big),$$

where the last inequality follows from the definition of height. In particular, for the given point $z \in C$, there is some index j such that

$$h\big(\xi_j(z)\big) \geq \frac{1}{n} h_{C,NA}(z) = \frac{N}{n} h_{C,A}(z) + O(1).$$

Reordering the coordinates, we may assume that $j = 1$. We similarly reorder the ξ_j''s to make $\xi_1'(w)$ have largest height. Combining this with (85) above gives the estimates

$$h\big(\xi_1(z)\big) \geq \frac{N}{ng}|z|^2 + O(1) \qquad \text{and} \qquad h\big(\xi_1'(w)\big) \geq \frac{N}{ng}|w|^2 + O(1). \tag{86}$$

In order to apply Roth's lemma, we need to use the section s to construct a polynomial. Recall that s is given by the collection of functions $\mathcal{F} = \{F_i(x, x')/y_i^d\}$. We consider the function field $K(C \times C)$ as a finite extension of the field $K(\xi_1, \xi_1')$ and compute the norm

$$Q(\xi_1, \xi_1') = \operatorname{Norm}_{K(C \times C)/K(\xi_1, \xi_1')} \frac{F_i(\xi, \xi')}{(y_i/y_0)^d}.$$

Note that Q is independent of i. Further, the function $F_i(\xi,\xi')(y_i/y_0)^{-d}$ is regular on the locus of $C \times C$ where $x_0 x_0' \neq 0$, so the function Q is a regular function on $\mathbb{A}^1 \times \mathbb{A}^1$. In other words, Q is in the polynomial ring $K[\xi_1, \xi_1']$.

From the definition of Q and the fact that F_i has degree δ_1 in x and δ_2 in x', we see that

$$\deg \xi_1(Q) \leq N d_1 \qquad \text{and} \qquad \deg \xi_2(Q) \leq N d_2. \tag{87}$$

Next, since Q is the product of N^2 terms, we have the estimate

$$\begin{aligned} h(Q) &\leq N^2 h(\mathfrak{F}) && \text{from (B.7.2)} \\ &\leq N^2 \left(c_{36} \frac{d_1 + d_2}{\gamma} + o(d_1 + d_2) \right) && \text{from (E.7.1)} \\ &\leq c_{37} d_1 / \gamma && \text{since } d_1 \geq d_2 \text{ by assumption.} \end{aligned} \tag{88}$$

We are going to apply Proposition E.11.1 with

$$P = Q, \quad r_1 = N d_1, \quad r_2 = N d_2, \quad \beta_1 = \xi_1(z), \quad \beta_2 = \xi_1'(w), \quad \omega = \varepsilon^2.$$

There are several things to be checked.

First, we have $r_2 \leq \omega r_1$, since this is just a restatement of the given condition $d_2 \leq \varepsilon^2 d_1$.

Second, we have

$$h(P) + 4r_1 = h(Q) + 4Nd_1 \leq c_{38} d_1 / \gamma$$

from (88). On the other hand, we can estimate

$$\begin{aligned} \omega r_1 h(\beta_1) &= \varepsilon^2 N d_1 h\big(\xi_1(z)\big) \\ &\geq \varepsilon^2 N d_1 \left(\frac{N}{ng} |z|^2 + O(1) \right) && \text{from (86)} \\ &\geq \frac{N^2 d_1}{ng} \left(\varepsilon^2 |z|^2 + O(1) \right) \\ &\geq c_{39} d_1 / \gamma && \text{from (84).} \end{aligned}$$

So if the constant in (84) is chosen sufficiently large, then the inequality

$$h(P) + 4r_1 \leq \omega r_1 h(\beta_1)$$

will be true; and a similar calculation again using (84) gives the corresponding inequality $h(P) + 4r_1 \leq \omega r_2 h(\beta_2)$

We are now in a position to apply Proposition E.11.1 to the polynomial Q, so we find indices i_1, i_2 satisfying

$$\frac{i_1}{r_1} + \frac{i_2}{r_2} \leq 4\sqrt{\omega} \qquad \text{and} \qquad (\partial_{i_1 i_2} Q)(\beta_1, \beta_2) \neq 0.$$

Using the definition of r_1, r_2, and ω, the first condition becomes

$$\frac{i_1}{d_1} + \frac{i_2}{d_2} \leq 4N\varepsilon.$$

As for the second condition, it certainly implies that (i_1, i_2) is an admissible pair for s, because when we use the product rule to differentiate Q, at least one of the terms in the resulting sum must be nonzero. This completes the proof of Proposition E.11.2. □

E.12. Completion of the Proof of Vojta's Inequality

We have now assembled all of the tools needed to prove Vojta's inequality (Theorem E.1.1), which we restate here for the convenience of the reader.

Theorem E.12.1. (Vojta's inequality) *There are two constants $\kappa_1 = \kappa(C)$ and $\kappa_2 = \kappa_2(g)$ such that if $z, w \in C(\bar{K})$ are two points satisfying*

$$|z| \geq \kappa_1 \qquad \text{and} \qquad |w| \geq \kappa_2 |z|, \tag{89}$$

then

$$\langle z, w\rangle \leq \tfrac{3}{4}|z|\,|w|.$$

Proof. We assume first that the constant κ_1 is chosen sufficiently large so that the finite set $Z \subset C$ described in Proposition E.10.1 contains no points with $|z| \geq \kappa_1$. Now suppose we are given two points $z, w \in C(\bar{K})$ satisfying (89), where we will be specifying κ_1 and κ_2 more precisely below. The reader is urged to verify at each stage that it is possible to choose κ_1 and κ_2 independently of $|z|$ and $|w|$.

Next we choose a large real number D and two small positive real numbers $1 > \varepsilon, \nu > 0$. In particular, we will assume that $D > |w|^2$. Eventually we will let $D \to \infty$, while we will choose values for ε and ν that depend only on C. We now set d_1, d_2, and d to have the values

$$d_1 = N\left[\sqrt{g+\nu}\frac{D}{|z|^2}\right], \quad d_2 = N\left[\sqrt{g+\nu}\frac{D}{|w|^2}\right], \quad d = N\left[\frac{D}{|z|\,|w|}\right], \tag{90}$$

and we consider the height of (z, w) relative to the Vojta divisor

$$\Omega = \Omega(d_1, d_2, d).$$

Note that our choice of d_1, d_2, d depends on the points z and w. The reader thus finally sees why it was always necessary in our estimates to

keep track of the dependence on d_1, d_2, and d. We are going to compare upper and lower bounds for $h_\Omega(z,w)$. These bounds depend on Ω, and since Ω depends on z and w via its dependence on d_1, d_2, d, we would end up with nothing if we did not have firm control over how all estimates depend on d_1, d_2, and d.

We begin with the lower bound provided by Proposition E.10.1, which says that

$$h_\Omega(z,w) \ge -h(\mathcal{F}) - c_{40}\big(i_1^*|z|^2 + i_2^*|w|^2\big) - c_{41}(i_1^* + i_2^* + \delta_1 + \delta_2 + 1).$$

This is valid for all pairs (i_1^*, i_2^*) that are admissible for the global section s to $\mathcal{O}(\Omega)$, where $\mathcal{F} = \{F_i\}$ is a set of polynomials corresponding to the section s. Assuming $\kappa_1 \ge 1$, we have $|z| \ge 1$ and $|w| \ge 1$, so adjusting the constants accordingly and using the definition of δ_1 and δ_2, we obtain the lower bound

$$h_\Omega(z,w) \ge -h(\mathcal{F}) - c_{42}\big(i_1^*|z|^2 + i_2^*|w|^2\big) - c_{43}(d_1 + d_2 + d). \tag{91}$$

Using our choice (90) for d_1, d_2, d and choosing κ_1 to satisfy $\kappa_1 > \varepsilon^{-1/2}$, we obtain

$$d_1 + d_2 + d \le ND\left(\frac{\sqrt{g+\nu}}{|z|^2} + \frac{2\sqrt{g+\nu}}{|w|^2} + \frac{1}{|z|\,|w|}\right) \le \frac{c_{44}D}{\kappa_1^2} \le c_{44}\varepsilon D. \tag{92}$$

Substituting (92) into (91) gives the lower bound

$$h_\Omega(z,w) \ge -h(\mathcal{F}) - c_{45}\big(i_1^*|z|^2 + i_2^*|w|^2\big) - c_{46}\varepsilon D. \tag{93}$$

Next we use Siegel's lemma (Proposition E.7.1) to find a "small" global section s to $\mathcal{O}(\Omega)$. In order to apply Proposition E.7.1 we need to verify that

$$d_1 d_2 - gd^2 \ge \gamma d_1 d_2 \tag{94}$$

for some positive constant γ that is independent of z and w. Using our choice (90) of d_1, d_2, d, we estimate

$$\begin{aligned}\frac{d_1 d_2 - gd^2}{d_1 d_2} &= 1 - \frac{gd^2}{d_1 d_2}\\ &\ge 1 - \frac{g\left(\dfrac{D}{|z||w|}\right)^2}{\left(\sqrt{g+\nu}\dfrac{D}{|z|^2} - 1\right)\left(\sqrt{g+\nu}\dfrac{D}{|w|^2} - 1\right)}\\ &= 1 - \frac{g}{g+\nu}\cdot\frac{1}{1 - \dfrac{|z|^2}{D\sqrt{g+\nu}}}\cdot\frac{1}{1 - \dfrac{|w|^2}{D\sqrt{g+\nu}}}.\end{aligned}$$

Hence as long as we fix a $\nu > 0$ and take D sufficiently large, the inequality (94) will be true with (say) $\gamma = \nu/3g$, and we can apply Proposition E.7.1. This means that we can find a section s given by a system of polynomials $\mathcal{F} = \{F_i\}$ satisfying

$$h(\mathcal{F}) \le c_{47}\frac{d_1+d_2}{\gamma} + o(d_1+d_2).$$

Using $\gamma = \nu/3g$ and the estimates (92), we obtain

$$h(\mathcal{F}) \le c_{48}(d_1+d_2) \le c_{49}\varepsilon D;$$

and substituting this into (93) gives us the lower bound

$$h_\Omega(z,w) \ge -c_{50}\left(i_1^*|z|^2 + i_2^*|w|^2\right) - c_{51}\varepsilon D. \tag{95}$$

Next we want to use Proposition E.11.2 to choose a small admissible pair (i_1^*, i_2^*). We need to verify the inequalities

$$\varepsilon^2 d_1 \ge d_2 \qquad \text{and} \qquad \min\left\{d_2|w|^2, d_1|z|^2\right\} \ge \frac{c_{52}}{\gamma\varepsilon^2}d_1 \tag{96}$$

required by Proposition E.11.2. To check the first one, we use (90) to compute

$$\frac{d_2}{d_1} \le \frac{N\sqrt{g+\nu}\,D/|w|^2}{N(\sqrt{g+\nu}\,D/|z|^2 - 1)} = \frac{2|z|^2}{|w|^2} \le \frac{2}{\kappa_2^2} \le \varepsilon^2,$$

provided that we pick $\kappa_2 \ge \sqrt{2}\varepsilon^{-1}$. Next we observe that $d_1|z|^2$ and $d_2|w|^2$ have comparable orders of magnitude, since there are numbers $0 \le \eta_1, \eta_2 < 1$ such that

$$\begin{aligned}\frac{d_2|w|^2}{d_1|z|^2} &= \left[\frac{\sqrt{g+\nu D}}{|w|^2}\right]|w|^2 \Big/ \left[\frac{\sqrt{g+\nu D}}{|z|^2}\right]|z|^2 \\ &= \left(1 - \frac{\eta_2|w|^2}{D\sqrt{g+\nu}}\right)\Big/\left(1 - \frac{\eta_1|z|^2}{D\sqrt{g+\nu}}\right).\end{aligned}$$

Thus, as soon as D is large enough, we obtain the estimate

$$\frac{1}{2} \le \frac{d_2|w|^2}{d_1|z|^2} \le 2.$$

Finally, the inequality in (96) will be true if we require κ_1 to satisfy $\kappa_1^2 \ge 2c_{52}/(\gamma\varepsilon^2)$, since we are given that $|z| \ge \kappa_1$.

We have now verified the conditions (96) needed to apply Proposition E.11.2, so we conclude that there is an admissible pair (i_1^*, i_2^*) for s satisfying

$$\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2} \le 4N\varepsilon.$$

In particular,

$$i_1^* \le 4N\varepsilon d_1 \qquad \text{and} \qquad i_2^* \le 4N\varepsilon d_2.$$

Substituting this into the lower bound (95) and absorbing the $4N$ into the constant gives

$$h_\Omega(z,w) \ge -c_{53}\varepsilon\big(d_1|z|^2 + d_2|w|^2\big) - c_{54}\varepsilon D. \tag{97}$$

Next we use (90) to estimate the quantities $d_1|z|^2$ and $d_2|w|^2$. Thus

$$d_1|z|^2 \le N\frac{\sqrt{g+\nu}D}{|z|^2}|z|^2 \le c_{55}D \quad \text{and} \quad d_2|w|^2 \le N\frac{\sqrt{g+\nu}D}{|w|^2}|w|^2 \le c_{56}D.$$

Substituting these into (97), we finally obtain the desired lower bound

$$h_\Omega(z,w) \ge -c_{57}\varepsilon D. \tag{98}$$

We are going to compare this with the upper bound given by Proposition E.4.1, which says that

$$h_\Omega(z,w) \le \frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle + c_{58}(d_1+d_2+d). \tag{99}$$

Using (92), we can replace the d_1+d_2+d by $c_{59}\varepsilon D$, and then combining the lower bound (98) with the upper bound (99) gives the inequality

$$\frac{d_1}{g}|z|^2 + \frac{d_2}{g}|w|^2 - 2d\langle z,w\rangle \ge -c_{60}\varepsilon D.$$

We next substitute the particular values (90) that we chose for d_1, d_2, d and divide by N to obtain

$$\frac{1}{g}\left[\sqrt{g+\nu}\frac{D}{|z|^2}\right]|z|^2 + \frac{1}{g}\left[\sqrt{g+\nu}\frac{D}{|w|^2}\right]|w|^2 - 2\left[\frac{D}{|z|\,|w|}\right]\langle z,w\rangle \ge -c_{61}\varepsilon D. \tag{100}$$

Note that (100) is true for all sufficiently large D. In particular, we can divide (100) by D and let D go to infinity. Noting that

$$\lim_{D\to\infty}\frac{1}{D}[\alpha D] = \alpha \qquad \text{for any real number } \alpha,$$

we obtain from (100) the inequality

$$2\frac{\sqrt{g+\nu}}{g} - 2\frac{\langle z,w\rangle}{|z|\,|w|} \ge -c_{61}\varepsilon.$$

A little bit of algebra then yields

$$\langle z,w\rangle \le \left(\frac{\sqrt{g+\nu}}{g} + \frac{1}{2}c_{61}\varepsilon\right)|z|\,|w|. \tag{101}$$

We are assuming that the genus satisfies $g \geq 2$, and so

$$\frac{1}{\sqrt{g}} \leq \frac{1}{\sqrt{2}} < \frac{3}{4}.$$

Hence if we choose ν and ε sufficiently small, then we will get

$$\langle z, w \rangle \leq \frac{3}{4}|z|\,|w|,$$

which is exactly Vojta's inequality. It is important to observe that we do not need to let $\varepsilon \to 0$. (This would force $\kappa_1, \kappa_2 \to \infty$, and then the statement of (E.12.1) would be vacuous.) The constant c_{61} appearing in (101) depends only on C (and the fixed embeddings ϕ_{NA} and ϕ_B), so it is possible to choose positive values for ε and ν that depend only on C, and that are independent of z and w. This completes the proof of Vojta's inequality (Theorem E.12.1), and in view of Proposition E.1.2, it also completes the proof of Faltings' theorem (E.0.1). □

Remark E.12.2. If the genus of C is large, then we can improve the constant appearing in Vojta's inequality. Thus fix any constant $1 > \varepsilon > 0$. Then the proof given above (see (101)) shows that there are constants κ_1, κ_2, depending only on C and ε, such that if $z, w \in C(\bar{K})$ are two points satisfying $|z| \geq \kappa_1$ and $|w| \geq \kappa_2|z|$, then

$$\langle z, w \rangle \leq \left(\frac{1}{\sqrt{g}} + \varepsilon\right)|z|\,|w|. \tag{102}$$

Remark E.12.3. Vojta's inequality alone is sufficient to imply the finiteness of $C(K)$. It is possible to give an effective upper bound for the *number* of points in $C(K)$ by combining Vojta's inequality with Mumford's gap principle B.6.6(a). See Exercise E.9.

Remark E.12.4. A useful remark of Oesterlé (unpublished) is that the constant κ_2 in Vojta's inequality (E.12.1) may be chosen purely in terms of the genus g of C (and on ε, of course). In fact, de Diego [1] has shown that κ_2 may even be chosen independent of g (cf. remark E.10.2). Although this may seem to be a minor point, it is very useful for deriving uniform upper bounds for the number of points in $C(K)$.

We conclude Part E with some brief remarks concerning certain special cases of Faltings' theorem (Mordell's conjecture) that can be proven by more elementary means. We are especially interested in cases for which one can effectively determine $C(K)$. Note that the proof of Faltings' theorem in this section might be called "semieffective," since it gives two effective constants $c_{1,\text{eff}}$ and $c_{2,\text{eff}}$, depending on C/K, with the property that the height of points in $C(K)$ is bounded by $c_{1,\text{eff}}$ with at most $c_{2,\text{eff}}$ exceptions.

This sort of statement is typical of results proven using Diophantine approximation techniques. Balanced against this positive result is the fact that at present, there is not a single curve C of genus $g \geq 2$ for which $C(K)$ can be effectively computed for all number fields K. Nevertheless, in very special cases (i.e., for special curves and special number fields) one can effectively determine the set of rational points. We list some instances where this can be done, proceeding from trivial to less trivial cases.

In the following remarks, C/K is a curve of genus $g \geq 2$ defined over a number field K as usual, and J is the Jacobian variety of C.

Remark E.12.5. If there exists a place v of K such that $C(K_v) = \emptyset$, then clearly $C(K) = \emptyset$. See Exercise E.13 for an example.

Remark E.12.6. If $J(K)$ is finite, that is, $J(K) = J(K)_{\text{tor}}$, then the set $C(K)$ can be effectively determined. Indeed, the torsion subgroup of $J(K)$ is always effectively computable and, after embedding C inside J, one can compute $C(K)$ as a subset of $J(K)$. See Exercise E.14 for an example.

More generally, if there is a nontrivial quotient B of J such that $B(K)$ is finite, then one can look at the composition $C \hookrightarrow J \rightarrow B$. The resulting map $C(K) \rightarrow B(K)$ is finite-to-one, and hence $C(K)$ is finite and can often be determined. See Exercise E.14 for an example. Note that this is the starting point of Mazur's proof (Mazur [1]) that modular curves $X_1(N)$ of positive genus have no noncuspidal rational points.

Remark E.12.7. If $J(K)$ has small rank, precisely, if $\operatorname{rank}(J(K)) \leq g-1$, Chabauty [1] already in 1941 gave a p-adic argument showing that $C(K)$ is finite. Although far from obvious, Coleman [2] has shown that Chabauty's proof can be refined to give a very strong effective bound for the number of points in $C(K)$. We record some special cases of Coleman's results.

Theorem E.12.7.1. (Coleman [2]) (i) *Let $C/\mathbb{Q}$ be a curve of genus 2 with good reduction at 2 or 3, and assume that* $\operatorname{rank} J(\mathbb{Q}) = 1$. *Then $\#C(\mathbb{Q}) \leq 12$. If in addition four of the Weierstrass points of C are rational and C has good reduction at 3, then $\#C(\mathbb{Q}) \leq 6$.*

(ii) *Let C/K be a curve of genus $g \geq 2$, let $p \geq 2g$ be a prime, let $\mathfrak{p}$ be an unramified prime of K lying over p, and suppose that C has good reduction at $\mathfrak{p}$. If* $\operatorname{rank} J(K) < g$, *then* $\#C(K) \leq N\mathfrak{p} + 2g\left(\sqrt{N\mathfrak{p}} + 1\right) - 1$.

In some cases Coleman's estimates are sufficiently sharp to allow the complete determination of $C(K)$. See Exercise E.15 for an example.

Remark E.12.8. Demjanenko [1] observed that if a curve C admits many independent maps to a single elliptic curve, then $C(K)$ is finite. Manin [5] generalized this result and gave an important application to rational points in towers of modular curves. We give the precise statement and sketch the proof.

Theorem E.12.8.1. (Demjanenko [1], Manin [5]) *Let C/K be a curve of genus $g \geq 2$, let B/K be an abelian variety, and let $f_1, \dots, f_r : C \to B$ be morphisms defined over K that are independent modulo constant morphisms. Assume further that* $\operatorname{rank} B(K) < r$. *Then $C(K)$ is finite. Further, it is possible to effectively bound the height of the points in $C(K)$ in terms of C, B, and the morphisms $f_1, \dots, f_r$. (See Exercise E.16 for an example.)*

PROOF. (sketch) Let $\operatorname{Mor}(C, B)$ denote the group of morphisms from C to B. Note that $\operatorname{Mor}(C, B)$ contains a subgroup isomorphic to B, namely the group of constant maps. Now fix an ample symmetric divisor D on B. A basic geometric fact says that the function

$$\operatorname{Mor}(C, B)/B \longrightarrow \mathbb{Z}, \qquad f \longmapsto \deg(f^*D),$$

is a positive quadratic form on the group $\operatorname{Mor}(C, B)/B$. We denote by $\langle \cdot, \cdot \rangle_{\deg}$ the associated bilinear form on $\operatorname{Mor}(C, B)/B$.

Next fix a divisor A of degree 1 on C. For any map $f : C \to B$, the divisor f^*D is algebraically equivalent to $\deg(f^*D)A$, so Theorem B.3.2 (or Theorem B.5.9) tells us that

$$\lim_{x \in C(\bar{K}),\, h_A(x) \to \infty} \frac{\hat{h}_D\big(f(x)\big)}{h_A(x)} = \deg(f^*D).$$

Applying this with f_i and $f_i + f_j$ and taking appropriate linear combinations, we obtain

$$\lim_{x \in C(\bar{K}),\, h_A(x) \to \infty} \frac{\langle f_i(x), f_j(x) \rangle_D}{h_A(x)} = \langle f_i, f_j \rangle_{\deg},$$

where $\langle \cdot, \cdot \rangle_D$ is the canonical height pairing on B with respect to the divisor D. Now taking the determinant over $1 \leq i, j \leq r$ yields

$$\lim_{\substack{x \in C(\bar{K}) \\ h_A(x) \to \infty}} \frac{\det\left[\big(\langle f_i(x), f_j(x) \rangle_D\big)_{1 \leq i,j \leq r}\right]}{h_A(x)^r} = \det\left[\big(\langle f_i, f_j \rangle_{\deg}\big)_{1 \leq i,j \leq r}\right] > 0,$$

where the positivity follows from the positive definiteness of the pairing $\langle \cdot, \cdot \rangle_{\deg}$ and the assumption that the maps $f_1, \dots, f_r : C \to B$ are independent.

It follows that $\det\left[\big(\langle f_i(x), f_j(x) \rangle_D\big)_{1 \leq i,j \leq r}\right]$ is positive, provided that $h_A(x)$ is sufficiently large . It follows from Proposition B.5.3 that the maps $f_1(x), \dots, f_r(x)$ are independent in $B(\bar{K}) \otimes \mathbb{R}$. We can rephrase this in a somewhat more illuminating way by saying that the set

$$\left\{x \in C(\bar{K}) \;\middle|\; f_1(x), \dots, f_r(x) \text{ are dependent in } B(\bar{K}) \otimes \mathbb{R}\right\}$$

is a set of bounded height. We also note that it is possible to compute an effective upper bound c_{eff}, depending only on the choice of the height functions h_A and h_D and the maps $f_1, \ldots, f_r$, such that every point x in this set satisfies $h_A(x) \le c_{\text{eff}}$.

Now we restrict to rational points $x \in C(K)$, and we use the assumption that $\operatorname{rank} B(K) < r$. It follows that $h_A(x) \le c_{\text{eff}}$, which completes the proof that $C(K)$ is an effectively determined finite set. □

We remark that Silverman [11] has generalized the result (E.12.8.1) of Demjanenko and Manin to the case of nonconstant families of abelian varieties (see also Lang [6, Chapter 12]).

Theorem E.12.8.2. (Silverman [11]) *Let C/K be a curve, let $B \to C$ be a family of abelian varieties defined over K, and let B' be the constant part of the family. (This means that almost every fiber is an abelian variety, and that B'/K is the "largest" constant abelian variety that embeds $B' \times C \hookrightarrow B$ over C.) For each $x \in C(\bar{K})$, let*

$$\sigma_x : \operatorname{Mor}(C, B) \longrightarrow B_x(\bar{K}), \qquad f \longrightarrow f(x),$$

be the specialization map. Then

$$\left\{x \in C(\bar{k}) \;\middle|\; \sigma_x : \operatorname{Mor}(C,B)/B'(\bar{k}) \to B_x(\bar{K})/B'_x(\bar{K}) \text{ is not injective}\right\}$$

is a set of bounded height. In particular, for all but finitely many $x \in C(k)$, the specialization map σ_x is injective on $\operatorname{Mor}_k(C, B)/B'(k)$.

Remark E.12.9. These examples do not cover all special cases or special methods for determining the rational points on curves. The most notable case of determining $C(K)$ via a highly indirect and difficult line of reasoning is surely Wiles's proof (Wiles [1]) of Fermat's Last "Theorem."

Exercises

E.1. Let V be a Euclidean vector space with inner product $\langle \cdot, \cdot \rangle$. For each point $x_0 \in V$ and each angle θ_0, let Γ_{x_0,θ_0} be the cone

$$\Gamma_{x_0,\theta_0} = \{x \in V \mid \theta(x, x_0) < \theta_0\}.$$

Prove that there are points $x_1, \ldots, x_n$ such that

$$n \le \left[2 + \frac{1}{\sin(\frac{1}{4}\cos^{-1}(\theta_0))}\right]^{\dim V} \qquad \text{and} \qquad V = \bigcup_{i=1}^{n} \Gamma_{x_i,\theta_0}.$$

E.2. Let C/K be a smooth curve of genus $g \geq 2$ defined over a number field K, and let J/K be the Jacobian of C.
(a) Use Vojta's inequality (E.1.1) to prove that there are constants $\kappa_1 = \kappa_1(C)$ and $\kappa_2 = \kappa_2(C)$ such that for every extension field L/K,

$$\#\{x \in C(L) \mid |x| \geq \kappa_1\} \leq \#J(L)_{\text{tors}} \cdot \kappa_2 \cdot 10^{\text{rank } J(L)}.$$

Note that κ_1 and κ_2 depend on C, but should be independent of the extension field L.
(b) Use Mumford's gap principle (B.6.6) to show that the term $\#J(L)_{\text{tors}}$ in (a) may be deleted.

E.3. (Liouville's inequality) Let S be any set of absolute values on a number field K.
(a) Prove that for all $\alpha \in K$ with $\alpha \neq 0$, one has the estimate

$$\sum_{v \in S} \log |\alpha|_v \geq -h([\alpha, 1]).$$

(b) If $K = \mathbb{Q}$ and $\alpha = a/b$, then describe the set S that one should choose such that the inequality in (a) is an equality.

E.4. Let C be a curve of genus $g \geq 2$ and $\mathcal{K}_C$ a canonical divisor on C.
(a) Prove that there is a divisor $A \in \operatorname{Div}(C)$ of degree 1 satisfying

$$(2g-2)A \sim \mathcal{K}_C.$$

(b) If in addition C is defined over a field K and $C(K)$ is not empty, prove that one can choose A such that the divisor class of A is defined over a field L/K with $[L:K] \leq (2g-2)^{2g}$, i.e., $A^\sigma \sim A$ for all $\sigma \in \operatorname{Gal}(L/K)$. To what extent can this upper bound for $[L:K]$ be improved?

E.5. Let C_1 and C_2 be smooth projective curves of genera g_1 and g_2, respectively. Recall that the arithmetic genus of a smooth projective surface S is defined by the formula $p_a(S) = h^2(S, \mathcal{O}_S) - h^1(S, \mathcal{O}_S)$. Prove that the arithmetic genus of the product $C_1 \times C_2$ is

$$p_a(C_1 \times C_2) = (g_1 - 1)(g_2 - 1) - 1.$$

E.6. Prove that the polynomial ring $K[\xi_1, \xi_2]$ contains $N\delta - \frac{1}{2}N(N-3)$ monomials of the form

$$\xi_1^i \xi_2^j \quad \text{with } 0 \leq j \leq N-1 \text{ and } 0 \leq i+j \leq \delta.$$

E.7. Let $p(\xi, \zeta) = \sum p_{ij}\zeta^i\xi^j$ and $\xi = \xi(\zeta)$ be as in Proposition E.9.1. Leibniz's formula says that

$$\partial_\ell p = \sum_{i,j} \sum_{\ell_0 + \cdots + \ell_i = \ell} p_{ij}(\partial_{\ell_0}\zeta^j)(\partial_{\ell_1}\xi)\cdots(\partial_{\ell_i}\xi).$$

Using this formula, the triangle inequality, and induction, give another proof of Proposition E.9.1 in the case that the absolute value v is nonarchimedean. (One can similarly prove a version of Proposition E.9.1 in the case that v is archimedean, but the resulting estimate would be too weak for our purposes.)

E.8. (Gelfand's inequality) Let $P_1, \dots, P_n \in \bar{\mathbb{Q}}[X_1, \dots, X_m]$ be polynomials in m variables with $\deg P_i \le d_i$. Then

$$h(P_1 \cdots P_n) \ge \sum_{i=1}^{n} h(P_i) - m(d_1 + \cdots + d_n).$$

E.9. Let C/K be a curve of genus $g \ge 2$ defined over a number field K, and let $x, z \in C(K)$.

(a) Mumford's gap principle says:

$$\text{If} \quad x \ne z \quad \text{and} \quad c_1 \le |x| \le |z| \le \beta|x| \quad \text{then} \quad \cos(x,z) \le \lambda;$$

and Vojta's inequality says:

$$\text{If} \quad c_2 \le |x| \quad \text{and} \quad \kappa|x| \le |z| \quad \text{then} \quad \cos(x,z) \le \lambda;$$

where $c_1, c_2, \beta, \kappa, \lambda$ are constants depending only on C/K, independent of x and z. Let $c_3 = \max\{c_1, c_2\}$. Combining Mumford's and Vojta's estimates, prove that there is a constant $N(\lambda)$ depending only on λ such that

$$\#\{x \in C(K) \mid |x| \ge c_3\} \le N(\lambda)\left(1 + \frac{\log \kappa}{\log \beta}\right).$$

(*Hint.* $N(\lambda)$ will be a bound for the number of cones necessary to cover $J(K) \otimes \mathbb{R}$, where each cone Γ has the property that every $x, y \in \Gamma$ satisfies $\cos(x,y) \le \lambda$.)

(b) Use (a), Exercise E.1, and Exercise B.10 to give an upper bound for $\#C(K)$.

E.10. As in the text, let A be a divisor of degree 1 on the projective curve C, let N be a sufficiently large integer so that NA is very ample, and identify C with its image $\phi_{NA}(C)$. Let $x_0, \dots, x_n$ be a basis of sections of $\mathcal{O}(NA)$. Let $U = \mathrm{GL}(n+1)$ be the variety of invertible matrices, and for $B = (b_{ij}) \in U$, write $x'_i = \sum_{j=0}^{n} b_{ij} x_j$.

(i) For $B \in U$ and distinct indices $i \ne j$, let L_{ij}^B denote the linear subspace $\{x'_i = x'_j = 0\}$. Prove that the set

$$\{B \in U \mid C \cap L_{ij}^B \ne \emptyset \text{ for some } i \ne j\}$$

is a proper Zariski subset of U.

(ii) Show that the set of B's such that $k(C) \ne k(x'_j/x'_i, x'_\ell/x'_i)$ is a proper Zariski subset of U. (*Hint.* Use the primitive element theorem, which says that if L/K is a finite separable extension and if L is generated by $\alpha_1, \dots, \alpha_m$, then $L = K(c_1\alpha_1 + \cdots + c_m\alpha_m)$ for all $(c_1, \dots, c_m) \in \mathbb{A}^m(K)$ lying outside a proper Zariski closed subset of $\mathbb{A}^m$. See, e.g., Lang [2, VII, 6, Theorem 14].)

E.11. It is clear that if C has genus $g \geq 2$, then Siegel's theorem (D.9.1) asserting that C has finitely many integral points is superseded by Faltings' theorem (E.0.1) saying that C has finitely many rational points. In this exercise you will show that Faltings' theorem can also be used to deduce finiteness of integral points on curves of genus 0 and 1 with an appropriate number of points at infinity.

(i) Let C be a smooth projective curve of genus 1, and let $U \subset C$ be an affine subset of C with at least one point at infinity. Prove that there exists a curve C' of genus $g' \geq 2$ and a covering $f : C' \to C$ such that f is unramified over U.

(ii) Assuming that the curves and maps in (i) are defined over a number field K, and letting S be a finite set of places of K, apply the Chevalley–Weil theorem (Exercise C.7) to prove that there exists a finite extension L/K such that $U(R_{K,S}) \subset f(C'(L))$. Deduce that $U(R_{K,S})$ is a finite set.

(iii) Repeat parts (i) and (ii) under the assumption that C has genus 0 and U has at least three points at infinity.

E.12. Let C be a curve of genus g, and let $A = (x_0)$ be an effective divisor of degree 1 on C. Find an explicit value of M such that the divisor

$$B := (M+1)(A \times C) + (M+1)(C \times A) - \Delta$$

is very ample on $C \times C$. (*Hint.* Show that $p_1^*\Theta + p_2^*\Theta + s_{12}^*\Theta$ is base-point free on $J \times J$, and that $3p_1^*\Theta + 3p_2^*\Theta$ is very ample on $J \times J$. Deduce that $4p_1^*\Theta + 4p_2^*\Theta + s_{12}^*\Theta$ is very ample on $J \times J$, and hence its pullback to $C \times C$ is very ample. Finally, show that the pullback has the form B and compute the value of M.)

E.13. Let $a \neq 0$ be an integer, let $n \geq 2$, and let C be the smooth plane curve defined by $X^n + Y^n + aZ^n = 0$.

(i) If $a > 0$ and n is even, prove that $C(\mathbb{Q}) = \emptyset$.

(ii) Let p be an odd prime such that $p | a$ and $p^n \nmid a$, and suppose that $\operatorname{ord}_2(n) \geq \operatorname{ord}_2(p-1)$, i.e., the highest power of 2 dividing $p-1$ also divides n. Prove that $C(\mathbb{Q}) = \emptyset$.

(*Hint.* Show that $C(\mathbb{R})$ or $C(\mathbb{Q}_p)$ is empty.)

E.14. Let C be the smooth projective curve with affine open subset U defined by $y^2 + y = x^5$, let $P_0 = (0,0)$, let $P_1 = (0,-1)$, and let P_∞ denote the point at infinity. Consider the Jacobian variety J of C and the natural embedding $j : C \to J$ defined by mapping P to the divisor class of $(P) - (P_\infty)$.

(i) It turns out that rank $J(\mathbb{Q}) = 0$. (You may try to prove this yourself, or see Fadeev [1].) Assuming this, prove that $J(\mathbb{Q}) \cong \mathbb{Z}/5\mathbb{Z}$.

(ii) Prove that $j(P_1) = 4j(P_0)$.

(iii) Prove that $2j(P_0), 3j(P_0) \notin j(C)$.

(iv) Conclude that $C(\mathbb{Q}) = \{P_0, P_1, P_\infty\}$.

(v) Use this exercise to prove Fermat's Last Theorem for exponent $p = 5$. (*Hint.* Use the fact that if $A^5 + D^5 = B^5$ with $D \neq 0$, then $(x, y) = (AB/D^2, A^5/D^5) \in C(\mathbb{Q})$.)

E.15. Let C be the hyperelliptic curve given by the affine equation

$$y^2 = x\left(x - \tfrac{1}{9}\right)(x^2 - 1)(x^2 - 18x + 1).$$

Note that C has two point P_∞ and P'_∞ at infinity. It turns out that the Jacobian variety of C satisfies $\operatorname{rank}\operatorname{Jac}(C)(\mathbb{Q}) = 1$. (You may try to prove this yourself, or see Flynn [2].) Assuming this fact, use Coleman's theorem (E.12.7.1) to prove that

$$C(\mathbb{Q}) = \left\{(0,0),(1,0),(-1,0),\left(\tfrac{1}{9},0\right),P_\infty,P'_\infty\right\}.$$

E.16. Let E/K be an elliptic curve with $\operatorname{rank} E(K) = 1$, and fix a Weierstrass equation $y^2 = x^3+ax+b = P(x)$ for E/K. Let C/K be a smooth projective model for the affine curve given by the equations

$$y^2 - P(x) = 0 \qquad \text{and} \qquad z^2 - P(x-d) = 0,$$

where we assume that $d \in K$ is chosen such that $P(X)$ and $P(X-d)$ have no common roots (this will be true for all but finitely many d's).

(i) Prove that C has genus 4.

(ii) Prove that there are two independent morphisms from C to E.

(iii) Use the argument of Demjanenko (E.12.8.1) to conclude that if $\operatorname{rank} E(K) = 1$, then $C(K)$ is finite.

PART F

Further Results and Open Problems

You can never plan the future by the past.
Edmund Burke, *Letter to the National Assembly*

We hope you have enjoyed our journey through the more or less tamed part of the world of Diophantine geometry. We now wish to take you on a last ride featuring some results whose proofs could not be included and a large number of open problems on the frontiers of knowledge. We describe conjectures and questions to serve as guideposts for future explorations into the relationships between arithmetic and geometry.

A simplified description of arithmetic geometry shows that it has developed along two paths, the Diophantine path and the modular path. The former includes the use of heights and Diophantine approximation, while the latter relates Diophantine problems to modular problems and relies on group representation techniques. It seems futile to attempt to predict which method will be most successful in the twenty-first century, the most likely guess being that a blend of the two will prove to be fruitful.

The present book is based on Diophantine methods. The original proof of Mordell's conjecture by Faltings (sketched below in Section F.4.2) is a mixture of Diophantine arguments and Galois representation theory. The subsequent proof of Vojta (with simplifications by Bombieri) described in Part E is purely Diophantine. In this context the introduction by Arakelov of an arithmetic intersection theory, imitating classical geometric intersection theory, has proven to be an invaluable insight, as advocated notably by Szpiro.

In a parallel development, mathematicians have linked elementary statements such as Fermat's last theorem and the *abc* conjecture to deep problems in Galois or automorphic representation theory, leading to the celebrated proof of the modularity conjecture and of Fermat's last theorem.

Our plan is to discuss first the generalization of Mordell's conjecture to higher-dimensional subvarieties of abelian varieties (Section F.1) and then to study the topology induced on these by the Néron–Tate norm (Section F.2). These two sections contain reasonably complete results. Next we consider conjectural upper and lower bounds for heights in various situations, including the famous *abc* conjecture of Masser and Oesterlé

(Section F.3). In the next section (F.4) we turn to the quest for effectivity, or more precisely, we explain why current methods give ineffective results. We also consider the question of quantitative results, that is, explicit upper bounds for the number of solutions to Diophantine problems. Finally, in the last section (F.5) we describe several far-reaching conjectures that describe the arithmetic properties of a variety in terms of its geometry. Here there are few general proven results, but many examples on which one can test conjectures.

We offer here a word of apology to those whose work is not quoted, due either to our arbitrary choices, our lack of competence, or a lack of space. By no means does this last part pretend to be a complete survey. For a discussion of many other results and conjectures, we refer the reader to Lang's volume of the *Russian Encyclopedia* (Lang [8]).

F.1. Curves and Abelian Varieties

Faltings' theorem (originally the Mordell conjecture) says that if X/k is a curve of genus at least 2 defined over a number field, then $X(k)$ is finite. There are several natural ways in which this statement may be generalized. For example, we might ask for which classes of varieties X/k is it true that $X(k)$ is finite. We will discuss this question later in Section F.4.2. Alternatively, we can view a curve of genus at least 1 as a subvariety of its Jacobian, so it is natural to study the Diophantine properties of higher-dimensional subvarieties of abelian varieties. Since the group of rational points of an abelian variety may be infinite, it is clear that if X/k is a subvariety of an abelian subvariety A/k, and if X contains an abelian subvariety B/k of A (or even a k-rational translate B), then $X(k)$ may be infinite. Lang conjectured that this is the only possible case, namely that $X(k)$ is, up to a finite set, the union of the rational points on translates of abelian subvarieties contained in X. Independently, Manin and Mumford asked whether it were true that a curve inside its Jacobian contains only a finite number of torsion points. Lang then proposed a very general conjecture (now a theorem) encompassing both questions on rational points and torsion points. We discuss this in the next section and then explain some applications to points of bounded degree on curves.

F.1.1. Rational Points on Subvarieties of Abelian Varieties

In order to state the main result of this section, we need one definition.

Definition. An abelian group Γ is said to have *finite rank* if it contains a free finitely generated subgroup $\Gamma_0 \subset \Gamma$ such that for every $x \in \Gamma$ there exists an integer $n \geq 1$ such that $nx \in \Gamma_0$.

The following conjecture of Lang (see Lang [1, 6, 12]) contains and generalizes both the Mordell and Manin–Mumford conjectures. It is now a theorem.

Theorem F.1.1.1. (née Lang's conjecture) *Let A be an abelian variety defined over $\mathbb{C}$, let X be a closed subvariety of A, and let Γ be a subgroup of $A(\mathbb{C})$ of finite rank. Then there exist a finite number of points $\gamma_1, \ldots, \gamma_r \in \Gamma$ and finite number of abelian subvarieties $B_1, \ldots, B_r$ of A such that $\gamma_i + B_i \subset X$ for all $1 \leq i \leq r$ and*

$$X(\mathbb{C}) \cap \Gamma = \bigcup_{1 \leq i \leq r} \gamma_i + (B_i(\mathbb{C}) \cap \Gamma).$$

In particular, if X does not contain any translate of a nontrivial abelian subvariety of A, then $X(\mathbb{C}) \cap \Gamma$ is finite.

For example, let X be a curve of genus at least 2 defined over a number field k, and (assuming $X(k) \neq \emptyset$) embed X into its Jacobian variety J. Taking $\Gamma = J(k)$, Theorem F.1.1.1 says that $X(k)$ is finite, which is Faltings' theorem. On the other hand, taking $\Gamma = J(\mathbb{C})_{\text{tors}}$, we find that $X(\mathbb{C})$ contains only finitely many torsion points of J, which is the Manin–Mumford conjecture. Thus Theorem F.1.1.1 provides a tremendous generalization of both Faltings' theorem and the Manin–Mumford conjecture.

Although the statement of (F.1.1.1) involves varieties and points defined over $\mathbb{C}$, we can start by selecting a finitely generated subgroup $\Gamma_0 \subset \Gamma$ such that every element of Γ has a multiple in Γ_0. If $\gamma_1, \ldots, \gamma_r$ generate Γ_0, then it is clear that $\gamma_1, \ldots, \gamma_r$ and the varieties X and A are all defined over a field K of finite type of $\mathbb{Q}$. Further, the fact that every $x \in \Gamma$ satisfies $nx \in G_0$ for some $n \geq 1$ shows that every point in Γ is algebraic over K, so is defined over $\bar{K}$. A specialization argument then allows one to reduce to the case that everything is defined over $\bar{\mathbb{Q}}$. Next, using Kummer-theoretic arguments (see below for references), it is possible to reduce to the case that Γ itself is finitely generated. This means that Γ is a subgroup of $A(k)$ for a number field k, so the problem is reduced to showing that $X(k)$ is contained in a finite union of sets of the form $\gamma + B(k)$, where B is an abelian subvariety of A (possibly $B = 0$), $\gamma \in A(k)$, and $\gamma + B \subset X$. We now give a sketch of the proof of this case. The original proof is due to Faltings (Faltings [2, 3]); detailed surveys may be found in the volume edited by Edixhoven and Evertse [1] and in the long paper of Vojta [4].

The proof follows the same pattern as the proof of Mordell's conjecture, albeit with many additional technical difficulties. As in Vojta's or Bombieri's proof, one constructs a line bundle with parameters, finds an upper bound for the canonical height of the rational points with respect to the chosen line bundle, and constructs a small section of the chosen line bundle in order to obtain a lower bound for the height.

In the proof of Mordell's conjecture we chose a linear combination of

$$p_1^*\Theta, \qquad p_2^*\Theta, \qquad \text{and} \qquad \mathcal{P} = s_{12}^*\Theta - p_1^*\Theta - p_2^*\Theta$$

(see Lemma E.2.1 for notation) and used its restriction to $X \times X$. By analogy, we select a symmetric ample line bundle $\mathcal{L}$ on A with associated Néron–Tate height $\hat{h}_{\mathcal{L}}(x) = \|x\|^2$ and consider linear combinations of the following line bundles on A^m (where just as in Roth's theorem, we will need to use a sufficiently large value for m):

$$\mathcal{L}_1 = p_1^*\mathcal{L}, \dots, \mathcal{L}_m = p_m^*\mathcal{L}, \qquad \text{and} \qquad \mathcal{P}_{ij} = (p_i + p_j)^*\mathcal{L} - p_i^*\mathcal{L} - p_j^*\mathcal{L}.$$

We will restrict these line bundles to X^m, or more generally, we will choose subvarieties X_i of X and restict attention to $X_1 \times \cdots \times X_m$. It is convenient to work with "bundles" in $\mathrm{Pic}_{\mathbb{Q}}$ (i.e., Pic tensored by $\mathbb{Q}$). Ampleness still makes sense in this context, but we may speak of sections only for suitable powers of the "bundle." The line bundle that is used for the proof of (F.1.1.1) has the form

$$\mathcal{L}(-\varepsilon, s) = \mathcal{L}(-\varepsilon, s_1, \dots, s_m) := \sum_{i=1}^{m-1} (s_i p_i - s_{i+1} p_{i+1})^*\mathcal{L} - \varepsilon \sum_{i=1}^{m} s_i^2 p_i^*\mathcal{L}$$

for certain rational numbers $s_1, \dots, s_m$. Since the s_i are only rational numbers, $s_i p_i - s_j p_j$ is not really a morphism, so $(s_i p_i - s_j p_j)^*\mathcal{L}$ is defined by noting that if the s_i's are integers, then

$$(s_i p_i - s_j p_j)^*\mathcal{L} = s_i^2 p_i^*\mathcal{L} + s_j^2 p_j^*\mathcal{L} - s_i s_j \mathcal{P}_{ij}.$$

The right-hand side is well-defined in $\mathrm{Pic}(A^m) \otimes \mathbb{Q}$ even when $s_i \in \mathbb{Q}$, which gives meaning to the left-hand side.

Remark F.1.1.2.

(i) If X is a curve and $m = 2$, then we recover the line bundle used in the proof of Mordell's conjecture: $\mathcal{L}(-\varepsilon, s_1, s_2) = d_1\mathcal{L}_1 + d_2\mathcal{L}_2 - d_3\mathcal{P}$ with $d_1 = s_1^2(1-\varepsilon)$, $d_2 = s_2^2(1-\varepsilon)$ and $d_3 = s_1 s_2$.

(ii) If X contains a nontrivial abelian subvariety of A, then $\mathcal{L}(-\varepsilon, s)$ is never ample on X^m (see Exercise F.1).

If m is large enough and ε small enough, if X is not a translate of an abelian subvariety, and if d is an integer chosen large enough and divisible by the denominators of the s_i's, then one can show (see Exercise F.2) that

$$h^0(X^m, \mathcal{L}(-\varepsilon, s)^{\otimes d}) \geq c(\varepsilon, m) d^{m \dim X} \prod_{i=1}^{m} s_i^{2 \dim X}.$$

Next, using basic properties of the Néron–Tate height (Theorem B.5.6), we can compute the height of a point $x = (x_1, \dots, x_m)$ with respect to the line bundle $\mathcal{L}(-\varepsilon, s)$ as

$$\hat{h}_{\mathcal{L}(-\varepsilon,s)}(x) = \sum_{i=1}^{m-1} s_i^2\|x_i\|^2 + s_{i+1}^2\|x_{i+1}\|^2 - 2s_i s_{i+1}\langle x_i, x_{i+1}\rangle - \varepsilon \sum_{i=1}^{m} s_i^2\|x_i\|^2.$$

Hence if we fix any $\eta > 0$ small enough and if we assume that the x_i's lie inside a small cone of $A(k) \otimes \mathbb{R}$, then we have

$$\langle x_i, x_j \rangle \geq (1-\eta)\|x_i\|\,\|x_j\|.$$

If we furthermore choose the s_i's to satisfy $s_i \sim s_1\|x_1\|/\|x_i\|$, then we get

$$\hat{h}_{\mathcal{L}(-\varepsilon,s)}(x_1,\ldots,x_m) \leq -ms_1^2\|x_1\|^2(\varepsilon-(n+1)\eta) \leq -\frac{m\varepsilon}{2}s_1^2\|x_1\|^2.$$

This gives the desired upper bound. In order to obtain a complementary lower bound, one proceeds as in the proof of Faltings' theorem by constructing a small section of $\mathcal{L}(-\varepsilon,s)^{\otimes d}$ and trying to find a derivative of that section that does not vanish at the given point. The higher-dimensional nonvanishing result needed was discovered by Faltings [2] (see also Edixhoven-Evertse [1]). We give here an explicit refinement worked out by Evertse [2].

Theorem F.1.1.3. (Product theorem, Faltings [2]) *Let $m \geq 2$, let $n = (n_1,\ldots,n_m)$ and $d = (d_1,\ldots,d_m)$ be m-tuples of positive integers, let $\sigma \geq 0$ and $0 < \varepsilon \leq 1$ be real numbers, and set $M = n_1 + \cdots + n_m$. We assume that these quantities satisfy*

$$\frac{d_h}{d_{h+1}} \geq \left(\frac{mM}{\varepsilon}\right)^M. \tag{$*$}$$

Let $\mathbb{P} = \mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_m}$. Let F be a multihomogeneous polynomial of degree d on $\mathbb{P}$, and let

$$Z_\sigma = \{x \in \mathbb{P} \mid \operatorname{ind}_d(F,x) \geq \sigma\}.$$

Suppose that Z_σ and $Z_{\sigma+\varepsilon}$ have a common irreducible component Z. Then there are subvarieties $Z_i \subset \mathbb{P}^{n_i}$ such that Z factors as the product $Z = Z_1 \times \cdots \times Z_m$.

Suppose further that F is defined over a field k_0. Let

$$s = \sum_{i=1}^m \operatorname{codim} Z_i \qquad \text{and} \qquad c_0 = 2\left(\frac{s}{\varepsilon}\right)^s m^M M^2.$$

Then the Z_i's are defined over a finite extension k_1/k_0 whose degree satisfies

$$[k_1 : k_0] \deg Z_1 \cdots \deg Z_m \leq \left(\frac{ms}{\varepsilon}\right)^s,$$

and the heights of the Z_i's satisfy

$$[k_1 : k_0] \deg Z_1 \cdots \deg Z_m \left(\sum_{i=1}^m \frac{d_i}{\deg Z_i} h(Z_i)\right) \leq c_0 \left(\sum_{i=1}^m d_i + h(F)\right).$$

(See Section F.2.2 below for the definition of the height $h(Z)$ of a variety. Intuitively, $h(Z)$ measures the height of the coefficients of the defining equations of Z.)

The product theorem plays a role analogous to that played by Roth's lemma in classical Diophantine approximation. Indeed, the product theorem can be used to prove Roth's theorem (see Evertse [2]) and even Schmidt's subspace theorem (F.5.3.4); see Faltings–Wüstholz [1]. The theorem can also be used on a product of projective varieties $X_1 \times \cdots \times X_m$ by using linear projections $\pi_i : X_i \to \mathbb{P}^{\dim X_i}$. This is similar to the argument in Part E, where we applied Roth's lemma to the product $C \times C$ by projecting it on $\mathbb{P}^1 \times \mathbb{P}^1$.

We will now suppose that we have points $x_1, \ldots, x_m$ with rapidly increasing heights:

$$\|x_1\| \geq C_0, \qquad \text{and} \qquad \|x_{i+1}\|/\|x_i\| \geq C_0.$$

Since $d_i/d_{i+1} = s_i^2/s_{i+1}^2 \sim \|x_{i+1}\|^2/\|x_i\|^2$, we can then apply the product theorem to the small section σ of $\mathcal{L}(-\varepsilon, s)$ using weights $d_i = ds_i^2$. The conclusion is that there exist subvarieties X_i of X such that $(x_1, \ldots, x_m) \in X_1 \times \cdots \times X_m$ and such that F vanishes on $X_1 \times \cdots \times X_m$. Since the degrees and heights of the X_i's are controlled, we can use induction to finish the proof of Theorem F.1.1.1. □

We close this section with a short bibliography and some further comments. The fundamental ideas of this section stem from Vojta's seminal paper (Vojta [1]), which provided a new proof of Mordell's conjecture. Vojta's methods are much closer to classical Diophantine approximation techniques than the methods of Faltings' original proof, although Vojta made extensive use of Arakelov theory. Faltings [2, 3] substantially simplified Vojta's proof, as he puts it "avoiding all the difficult Arakelov theory," and moreover extended it to higher-dimensional varieties, the main new tool being the product theorem. Earlier, Raynaud [1, 2, 3] had proven the generalized Manin–Mumford conjecture, that is, Lang's conjecture with $\Gamma = A_{\text{tor}}$. A different proof of the Manin–Mumford conjecture for curves was given by Coleman [1] using p-adic abelian integrals. Coleman [2] also observed that in very special cases, his theory, combined with an old argument of Chabauty, yields effective finiteness of rational points on curves (see notes at the end of Part E). Another proof of the generalized Manin–Mumford conjecture was proposed by Hindry [1], following a suggestion of Serge Lang for curves and relying on a difficult result of Serre on Galois representations associated to torsion points on abelian varieties. The same paper also contains the reduction of the general Lang conjecture (subgroup of finite rank in $A(\mathbb{C})$) to the fundamental case treated by Faltings (where Γ is the Mordell–Weil group $A(k)$, hence is finitely generated).

Though in some sense the "arithmetic" case lies deeper, the function field case is also of great interest, with extra difficulties in characteristic p. In that case one is led to formulate a relative Lang conjecture, which is also now a theorem. Manin [1] was the first to prove Mordell's conjecture over function fields (in characteristic 0). Buium, in a series of papers (Buium [1, 2, 3]), has given a new approach and generalized Manin's work to higher-dimensional varieties. He also obtains bounds for the cardinality of the intersection $\Gamma \cap X$ that depends only on rank(Γ) and on the degree of a polarization on A and of X (with respect to this polarization). Abramovic and Voloch [1] have given a proof of a version of Lang's conjecture in characteristic p under additional assumptions. Hrushovski [1] has found a model-theoretic proof that covers both characteristic 0 and characteristic p. The proof starts, roughly speaking, like Buium's theory, but then makes heavy use of classification results in model theory by Hrushovski himself and Zilber (see Bouscaren [1] for a detailed survey or Poizat's Bourbaki talk, Poizat [1]). It came as a surprise to many number theorists that a branch of mathematical logic could have such arithmetical applications.

Finally, the conjecture of Lang can also be formulated (and proved) for subvarieties of semiabelian varieties, that is, extensions of abelian varieties by tori. Hrushovski's methods work also for semiabelian varieties over function fields of any characteristic. Vojta [2] has proven the analogue of Faltings' theorem for semiabelian varieties over number fields by extending his and Faltings' methods. The analogue of the reduction from subgroups of finite rank to subgroups of finite type is due to McQuillan [1], extending the methods of Hindry. The Manin–Mumford conjecture can even be extended to subvarieties of any commutative algebraic group (Hindry [1]). We also mention that the general case of subvarieties of $\mathbb{G}_m^s$ was treated earlier by Laurent [1], with partial cases treated even earlier by Liardet [1].

F.1.2. Application to Points of Bounded Degree on Curves

Let X be a curve of genus g defined over a number field k. If $g \geq 2$ and if K/k is any fixed finite extension of k, then we know that $X(K)$ is finite. In this section we consider what happens when the field K is allowed to vary. More precisely, we ask:

$$\text{When is} \quad X^{(d)}(k) := \bigcup_{[K:k] \leq d} X(K) \quad \text{finite?}$$

As the next result makes clear, the answer is closely connected to Lang's conjecture (Theorem F.1.1.1).

Theorem F.1.2.1. *Let X be a curve of genus $g \geq 2$, let $d \geq 1$ be an integer, and let $W_d(X) := X + \cdots + X \subset \operatorname{Jac}(X)$. If X admits no*

morphisms $X \to \mathbb{P}^1$ *of degree less than or equal to* d *and if* $W_d(X)$ *contains no translates of abelian subvarieties of* $\mathrm{Jac}(X)$, *then the set*

$$X^{(d)}(k) := \bigcup_{[K:k]\le d} X(K)$$

is finite. In other words, $X(\bar{k})$ *contains only finitely many points defined over fields of degree less than or equal to* d.

Further, if either of the hypotheses is false, then there exists a finite extension k' *of* k *such that* $X^{(d)}(k')$ *is infinite.*

PROOF. For simplicity we will assume that $X(k) \neq \emptyset$ and leave the reader to fill in the details otherwise. This allows us to construct an embedding $j : X \to \mathrm{Jac}(X)$ defined over k.

A point of degree $r \le d$ over k on X corresponds to a k-irreducible effective divisor of degree r, hence to a point in the r-fold symmetric product $\mathrm{Sym}^r(X)(k)$. Fixing $P_0 \in X(k)$, we get an injection $D \mapsto D + (d-r)(P_0)$ from $\mathrm{Sym}^r(X)(k)$ into $\mathrm{Sym}^d(X)(k)$. In particular $X^{(d)}(k)$ is finite if and only if $\mathrm{Sym}^d(X)(k)$ is finite.

Next consider the natural map

$$\Phi_d : \mathrm{Sym}^d(X) \to W_d(X) \subset \mathrm{Jac}(X), \quad \{x_1, \ldots, x_d\} \mapsto j(x_1) + \cdots + j(x_d).$$

The Abel–Jacobi theorem tells us that the map Φ_d is injective if and only if there is no linear system of degree at most d and dimension al least 1 on X. Such a linear system corresponds to a map of degree at most d to $\mathbb{P}^1$. Thus under our hypotheses, $\mathrm{Sym}^d(X)(k)$ is finite if and only if $W_d(X)(k)$ is finite. Now Theorem F.1.1.1 tells us that $W_d(X)(k)$ is finite, provided that $W_d(X)$ contains no abelian subvarieties. This completes the proof of the main part of the theorem. The final remarks are clear from the proof. □

In view of Theorem F.1.2.1, we would like to know when a curve X is likely to admit a morphism of degree at most d to $\mathbb{P}^1$. From Section A.3, a morphism of degree d to $\mathbb{P}^1$ corresponds to a linear system of dimension 1 and degree d. In the literature, such a linear system is called a g^1_d. It is customary to say that a curve is *d-gonal* if it admits a morphism of degree d to $\mathbb{P}^1$. Special cases include hyperelliptic curves (2-gonal) and trigonal curves (3-gonal).

Theorem F.1.2.2. (Existence of morphisms to $\mathbb{P}^1$) *Let* X *be a curve of genus* g.

(i) *If* $d \ge g/2+1$, *then there exists a nonconstant morphism* $X \to \mathbb{P}^1$ *of degree less than or equal to* d.

(ii) *If $d < g/2 + 1$ and X is "general" (see below for the definition), then there are no nonconstant morphisms $X \to \mathbb{P}^1$ of degree less than or equal to d.*

PROOF. (i) See Arbarelo–Clemens–Griffiths–Harris [1, Chapter 5, Existence Theorem 1.1, page 206].
(ii) See Arbarelo–Clemens–Griffiths–Harris [1, Dimension Theorem 1.5, page 214]. □

The word "general" in the statement of (F.1.2.2ii) means the following. Isomorphism classes of curves of genus g are parametrized by a quasi-projective moduli variety $\mathcal{M}_g$. Then there exist a countable number $Z_1, Z_2, \ldots$ of proper Zariski closed subvarieties of $\mathcal{M}_g$ such that (F.1.2.2ii) holds for all curves X whose isomorphism class $[X]$ lies in $\mathcal{M}_g \smallsetminus \cup_{n\geq 1} Z_n$. This is satisfactory from the viewpoint of algebraic geometry, since the complement of the countably many Z_n's is a large set due to the uncountability of $\mathbb{C}$. However, it is much less satisfactory from an arithmetic viewpoint, since it is conceivable (although unlikely) that the union of the Z_n's could contain $\mathcal{M}_g(\bar{\mathbb{Q}})$. This possibility cannot be ruled out purely on the basis of cardinality, since $\bar{\mathbb{Q}}$ is countable.

Since it is known that the general Jacobian is a simple abelian variety, we obtain the following.

Corollary F.1.2.3. *Let X be a general curve of genus g, let $d < g/2 + 1$, and let k be any finitely generated field k over which X is defined. Then $X^{(d)}(k)$.*

It is harder to determine when W_d contains an abelian subvariety of $\mathrm{Jac}(X)$, and this problem gives rise to some questions of geometric interest. To start, we observe that it is easy to construct curves whose associated W_d's contain abelian varieties. For example, if $f : X \to E$ is a map of degree d from X to an elliptic curve, then $f^* : E \to \mathrm{Sym}^d(X)$, and hence E sits inside W_d. More generally, a morphism $f : X \to Y$ of degree d from X to a curve Y of genus $h \geq 1$ induces a map $f^* : \mathrm{Sym}^h(Y) \to \mathrm{Sym}^{dh}(X)$, and this gives a copy of $\mathrm{Jac}(Y)$ inside $W_{hd}(X)$. Thus $W_{hd}(X)$ contains an abelian variety of dimension h.

Extending this idea further, suppose that there are covering maps

$$\begin{array}{ccccc} & & Z & & \\ & \swarrow^{f} & & \searrow^{p} & \\ X & & & & Y \end{array} \qquad \begin{array}{l} \deg(p) = e, \\ \mathrm{genus}(Y) = h. \end{array}$$

Then we get maps

$$\mathrm{Sym}^h(Y) \xrightarrow{p^*} \mathrm{Sym}^{eh}(Z) \xrightarrow{f_*} \mathrm{Sym}^{eh}(X),$$

where $f_*\left(\sum n_i(x_i)\right) = \sum n_i\big(f(x_i)\big)$. The composition $f_* \circ p^*$ induces a finite map $\mathrm{Jac}(Y) \to W_{eh}(X)$, so $W_{eh}(X)$ contains an abelian variety of dimension h.

Examples. (1) We construct curves X of genus $g \geq 3$ that are *bielliptic*, that is, are double covers of an elliptic curve E. It is clear that if X is such a curve, then $W_2(X)$ will contain a copy of E. Let E and X be the curves

$$E : y^2 = P(x) = x^3 + ax + b, \qquad X : y^4 = P(x) = x^3 + ax + b.$$

The map $(x, y) \mapsto (x, y^2)$ is clearly a double cover $X \to E$, and it is easy to check that X has genus 3.
(2) More generally, let $Q(x)$ be a polynomial of degree m such that the product $P(x)Q(x)$ has no double roots, and let X be a smooth projective curve birational to the curve in $\mathbb{A}^3$ given by the equations $y^2 = P(x)$ and $z^2 = Q(x)$. Then $(x, y, z) \mapsto (x, y)$ exhibits X as a double cover of E, and X has genus $m + 1$. See Exercise F.3 for examples of curves of genus g whose W_{2h} contains an abelian subvariety of dimension h when $2h \leq g - 2$.
(3) For an example in the positive direction, consider the curves $X = X_0(p)$ that parametrize elliptic curves having a cyclic subgroup of order p. One can show that if $p > 61$ and $p \neq 71, 79, 83, 89, 101$, or 131, then $X^{(2)}(k)$ is finite for every number field k. Further, all of the excluded p's correspond to curves that have genus less than 2 or are hyperelliptic or are bielliptic. See Frey [4], Hindry [3], and Harris-Silverman [1] for details.

We next quote two general results.

Theorem F.1.2.4. *Let X be a curve of genus g.*
(i) (Debarre–Fahlaoui [1]) *If $d < g$ and if $W_d(X)$ contains an abelian variety A, then* $\dim A \leq d/2$.
(ii) (Frey [3]) *If X is defined over a number field k and if $X^{(d)}(k)$ is infinite, then there exists a morphism $f : X \to \mathbb{P}^1$ defined over k of degree less than or equal to $2d$.*

The examples given above show that (i) is sharp. The next result deals with small values of d. (See also Harris–Silverman [1] and Hindry [3].)

Theorem F.1.2.5. (Abramovic–Harris [1]) *Let X be a curve of genus g.*
(i) *If* $\operatorname{Sym}^2 X$ *contains an elliptic curve, then X is bielliptic; and further, if $g \geq 4$, then X is not hyperelliptic.*
(ii) *If W_3 contains an elliptic curve E and $g \geq 5$, then there is a covering $f : X \to E$ with* $\deg(f) \leq 3$; *and further, if $g \geq 8$, then X is not trigonal.*
(iii) *If W_4 contains an elliptic curve E and $g \geq 8$, then either there is a covering $X \to E$ of degree at most 4 or else there is a covering $X \to X'$ of degree 2 with* $\operatorname{genus}(X') = 2$.
(iv) *If W_4 contains an abelian variety A of dimension 2 and $g \geq 6$, then there is a covering $X \to X'$ of degree 2 with* $\operatorname{genus}(X') = 2$.

An interesting example is provided by the genus-3 curve $y^2 = x^8 + 1$, which is both hyperelliptic and a double cover of the elliptic curve $y^2 = x^4 + 1$. Another interesting example is the genus-7 curve $y^3 = x^9 + 1$, which is both trigonal and a triple cover of the elliptic curve $y^3 = x^3 + 1$.

Debarre and Fahlaoui [1] show that if W_{2h} contains an h-dimensional abelian variety and if $h < g/3$, then there is a covering $X \to X'$ of degree 2 with genus$(X') = h$. They also prove that if W_d contains an h-dimensional abelian variety with $d < g/6$ (or $d(d-1)+2 < 2g$) and $h > d/4$, then there is a covering $X \to X'$ of degree 2 or 3 with genus$(X') = h$.

For (smooth) plane curves there is the following nice result.

Theorem F.1.2.6. (Debarre–Klassen [1]) *Let X be a smooth plane curve of degree d, and thus of genus $g = (d-1)(d-2)/2$, defined over a number field k.*
(a) *If $d \geq 7$, then $X^{(d-2)}(k)$ is finite.*
(b) *If $d = 4$, 5, or 6, then $X^{(d-2)}(k)$ is finite if X does not admit a morphism $X \to E$ of degree $d-2$ to an elliptic curve E.*

Notice that (F.1.2.6) is essentially optimal, since if $P_0 \in X(k)$ is any rational point, then the lines defined over k and passing through P_0 will intersect X at points of degree at most $d-1$, hence will lead to infinitely many points in $X^{(d-1)}(k)$.

The restriction to $d \geq 7$ in (F.1.2.6a) is also necessary, since the Fermat sextic X given by the equation $x^6+y^6=z^6$ is clearly a cover of degree 4 of the elliptic curve $u^3+v^3=w^3$, and hence $W_4(X)$ contains an elliptic curve. Similarly, the Fermat quartic X given by the equation $x^4+y^4=z^4$ is a degree-two cover of the elliptic curve given (in affine form) by $v^4+1=u^2$, and hence $W_2(X)$ contains an elliptic curve.

F.2. Discreteness of Algebraic Points

Once it is known (or conjectured) that certain sets of points of bounded degree are finite, it is natural to consider Diophantine approximation questions as the size of the set is allowed to grow. For example, it is known that there are only finitely many points of bounded degree and canonical height on an abelian variety, and further that the points of height 0 are precisely the torsion points. So we might ask:

> Does there exist a sequence of *nontorsion* points
> $P_1, P_2, \ldots \in A(\bar{k})$ such that $\lim_{n\to\infty} \hat{h}(P_n) = 0$?

The answer to this question is clearly yes. We simply take any nontorsion point $P_1 \in A(\bar{k})$, and then for each $n \geq 1$ we choose a point $P_n \in A(\bar{k})$ satisfying $[n]P_n = P_1$. Then

$$\hat{h}(P_n) = \frac{1}{n^2}\hat{h}(P_1) \xrightarrow[n\to\infty]{} 0.$$

Bogomolov suggested that this should be essentially the only way to get a positive answer to our question. In particular, he conjectured that

if the P_n's are required to lie on a curve X of genus at least 2, or more generally on a subvariety X of A that does not contain a translate of an abelian subvariety, then there cannot exist a sequence of distinct points in $X(\bar{k})$ whose height tends to 0. Notice that this strengthens Raynaud's theorem [1, 2, 3] (Manin–Mumford conjecture), since it implies in particular that X cannot contain infinitely many torsion points.

This sort of question is closely connected with the arithmetic complexity of the variety X. To make this idea precise, we need to define the height of the variety X. We now discuss Bogomolov's conjecture, followed by a brief introduction to the theory of heights of subvarieties and cycles.

F.2.1. Bogomolov's Conjecture

The following refinement of Raynaud's theorem (Manin–Mumford conjecture) was conjectured by Bogomolov [1] and proven by Ullmo.

Theorem F.2.1.1. (Ullmo [1]) *Let $X/\bar{\mathbb{Q}}$ be a curve of genus $g \geq 2$ sitting in its Jacobian J, and let $\|\cdot\| = \sqrt{\hat{h}(\cdot)}$ be the seminorm on $J(\bar{\mathbb{Q}})$ provided by the Néron–Tate height relative to an ample symmetric divisor on J. Then the topology on $X(\bar{\mathbb{Q}})$ induced by the seminorm $\|\cdot\|$ is discrete. In other words, for every $P \in X(\bar{\mathbb{Q}})$ there exists an $\varepsilon > 0$ such that the set*

$$\{Q \in X(\bar{\mathbb{Q}}) \mid \|P - Q\| < \varepsilon\}$$

is finite.

Before Ullmo's proof of (F.2.1.1), special cases were known through the work of Szpiro [4], Zhang [1], and Burnol [1]. A generalization to subvarieties of higher dimension has also been proven. Before giving the result, we need one definition. A *torsion subvariety* of an abelian variety A is a subvariety of the form $b + B$, where b is a torsion point of A, and B is an abelian subvariety of A. For example, a torsion point is automatically a torsion subvariety, and if A is simple (i.e., has no nontrivial abelian subvarieties), then these are the only torsion subvarieties of A.

Theorem F.2.1.2. (Ullmo–Zhang, Zhang [2]) *Let $X/\bar{\mathbb{Q}}$ be a subvariety of an abelian variety $A/\bar{\mathbb{Q}}$, let Z be the union of all torsion subvarieties of X, and let $U := X \smallsetminus Z$. Let $\|\cdot\| = \sqrt{\hat{h}(\cdot)}$ be the seminorm on $A(\bar{\mathbb{Q}})$ provided by the Néron–Tate height relative to an ample symmetric divisor on A. Then $\|\cdot\|$ induces the discrete topology on $U(\bar{\mathbb{Q}})$. More precisely, for all $P \in X(\bar{\mathbb{Q}})$ there exists an $\varepsilon > 0$ such that the set*

$$\{Q \in U(\bar{\mathbb{Q}}) \mid \|P - Q\| < \varepsilon\}$$

is finite.

We will not say anything about the proof of (F.2.1.2) except that it relies heavily on Arakelovian methods and properties of equidistribution of

points of small height proved by Szpiro–Ullmo–Zhang [1]. A second proof was found by by David–Philippon [1], which is more elementary in the sense that it does not use Arakelov theory.

The multiplicative analogue of (F.2.1.2), that is, the analogous result for subvarieties of $(\mathbb{G}_m)^s$, is also known and was first proved by Zhang [3] as an application of his arithmetical ampleness results. This analogue states that if X is a subvariety of $(\mathbb{G}_m)^s$ and if U is the Zariski open subset obtained by deleting torsion subvarieties from X, then the Weil height induces a discrete topology on $U(\bar{k})$. A more elementary proof was given by Schmidt [3] and then extended by Bombieri–Zannier [1]. The case of abelian varieties with complex multiplication was settled by Bombieri–Zannier [2] using a similar method. Bilu [1] later provided a proof of the equidistribution property of points of small heights, thus giving another approach to the analogue of Bogomolov's conjecture.

It is tempting to try to merge Bogomolov and Lang conjectures into a single result. This has been done by Poonen.

Theorem F.2.1.3. (Poonen [1]) *Let A be an abelian variety defined over a number field k, let Γ be a subgroup of finite rank in $A(\bar{k})$, and for any $\varepsilon > 0$, define*

$$\Gamma_\varepsilon := \{\gamma + z \mid \gamma \in \Gamma,\ z \in A(\bar{k}),\ \text{and } \hat{h}(z) \leq \varepsilon\}.$$

Let X be a closed subvariety of A that is not equal to the translate of an abelian subvariety of A. Then there exists an $\varepsilon > 0$ (depending on A, X, and Γ) such that $X(\bar{k}) \cap \Gamma_\varepsilon$ is not Zariski dense in X.

Poonen also shows how to extend (F.2.1.3) to semiabelian varieties, provided that the analogue of Bogomolov's conjecture and equidistribution of small points is true. The latter has now been proven for a group variety isogenous to the product of an abelian variety by a linear torus (Chambert-Loir [1]).

F.2.2. The Height of a Variety

The height of a point is a measure of its arithmetic complexity. A point is simply a variety of dimension 0, so it is natural to look for a way to measure the arithmetic complexity of higher-dimensional varieties. For example, if X is a hypersurface of degree d in $\mathbb{P}^n$, then X is defined by a single homogeneous equation

$$F(x) = \sum_{i_0+\cdots+i_n=d} a_{i_0\ldots i_d} x_0^{i_0} \cdots x_d^{i_d} = 0,$$

and F is uniquely determined by X up to multiplication by a nonzero constant. It is then natural to define

$$h(X) = h(F) = h(a),$$

where a is the point in projective space whose homogeneous coordinates are the coefficients $a_{i_0 \dots i_d}$ of F.

This construction may be generalized from hypersurfaces to arbitrary subvarieties of $\mathbb{P}^n$ by using Chow forms. Recall (see Shafarevich [1, Chapter I.6.5] or Exercise A.1.17) that to each variety X of degree d and dimension r in $\mathbb{P}^n$ there is associated a multihomogeneous form F_X of multidegree $(d, \dots, d)$ determined by the property that

$$F_X(a_0^{(0)}, \dots, a_n^{(0)}; a_0^{(1)}, \dots, a_n^{(1)}; \dots ; a_0^{(r)}, \dots, a_n^{(r)}) = 0$$

if and only if X has a nonempty intersection with each of the $r+1$ hyperplanes

$$a_0^{(i)} X_0 + \cdots + a_n^{(i)} X_n = 0, \qquad 0 \le i \le r.$$

The form F_X is the *Chow form* (also called the *Cayley form*) of X. We then define the height of X by

$$h(X) = h(F_X);$$

that is, $h(X)$ is the height of the point whose projective coordinates are the coefficients of its Chow form F_X. More generally, if X is a variety and if D is a very ample divisor on X, we fix an associated embedding $\phi_D : X \hookrightarrow \mathbb{P}^n$ and define the height of a subvariety $X \subset X$ relative to D to be $h_D(X) := h(\phi_D(X))$. Of course, the value of $h_D(X)$ depends on the embedding, but only up to the usual $O(\deg X)$.

An alternative to the use of Chow coordinates is the arithmetic intersection theory developed by Arakelov–Gillet–Soulé. A conceptual insight provided by this approach is the analogy between the projective degree of a variety and its height. Thus if $X/\mathbb{Q}$ is a variety of dimension r and if $\mathcal{L}$ an (ample) line bundle on X, the projective degree of X with respect to $\mathcal{L}$ is

$$\deg_{\mathcal{L}} X = \deg_0(X \cdot \mathcal{L}^r).$$

Now choose a projective model $f : \mathfrak{X} \to \operatorname{Spec}(\mathbb{Z})$ for X over $\operatorname{Spec}(\mathbb{Z})$ and extend $\mathcal{L}$ to a line bundle on $\mathfrak{X}$. Further, choose archimedean metrics for the fibers of $\mathcal{L}$ over the archimedean places, and denote the resulting metrized line bundle by $\bar{\mathcal{L}}$. Then one can define an arithmetic intersection $\mathfrak{X} \cdot \bar{\mathcal{L}}^{r+1}$, take its pushdown via f, and compute the Arakelov degree to define (following Faltings [2] and Bost–Gillet–Soulé [1])

$$h_{\bar{\mathcal{L}}}(X) = \deg_{\mathrm{Ar}} f_* \left(\mathfrak{X} \cdot \bar{\mathcal{L}}^{r+1} \right).$$

We also note that this height is sometimes normalized by dividing it by the projective degree $\deg_{\mathcal{L}} X$. We refer the reader to Bost–Gillet–Soulé [1] and Gubler [1] for further developments on arithmetic intersection theory and heights, such as an arithmetic Bézout's theorem.

We next introduce some additional ways of defining the height of a curve or an abelian variety. We begin by formalizing the idea that the collection of all curves of genus g fits into some sort of universal family, and similarly for the collection of all (principally polarized) abelian varieties of dimension g.

Definition. (Mumford) Let M be a collection of (isomorphism classes) of algebraic varieties, possibly with some additional structure. For example, M might be the collection of all of smooth projective curves of genus g, or M might be the collection of all principally polarized abelian varieties with a full level-N structure (see below).

(i) A *coarse moduli space for* M is a variety $\mathcal{M}$ such that if $f : \mathcal{X} \to T$ is a family of elements of M (i.e., each fiber $\mathcal{X}_t = f^{-1}(t)$ is in M), then there is a morphism $\Phi : T \to \mathcal{M}$ with the property that

$$\Phi(s) = \Phi(t) \text{ if and only if } \mathcal{X}_s \cong \mathcal{X}_t.$$

(ii) A *fine moduli space for* M is a variety $\mathcal{M}$ together with a universal family $\pi : \mathcal{U} \to \mathcal{M}$ such that if $f : \mathcal{X} \to T$ is a family of elements of M as above, then there is a morphism $\Phi : T \to \mathcal{M}$ such that $\mathcal{X}$ is the pullback of $\mathcal{U}$ via Φ and with the property that

$$\Phi(s) = \Phi(t) \text{ if and only if } \mathcal{U}_{\Phi(s)} \cong \mathcal{U}_{\Phi(t)}.$$

It is rare for a fine moduli space to exist for a class of varieties without the introduction of some additional structure. We illustrate this idea by looking at principally polarized abelian varieties with *full level-N structure.* This means that we classify triples $(A, \lambda, \varepsilon)$, where A is an abelian variety, λ is a principal polarization of A, and ε is a fixed isomorphism

$$\varepsilon : (\mathbb{Z}/N\mathbb{Z})^{2g} \longrightarrow A[N].$$

Two triples $(A, \lambda, \varepsilon)$ and $(A', \lambda', \varepsilon')$ are isomorphic if there is an isomorphism of abelian varieties $\alpha : A \to A'$ such that $\alpha^*(\lambda') = \lambda$ and $\varepsilon' = \alpha \circ \varepsilon$. Notice, for example, that every principally polarized abelian variety has a nontrivial automorphism, namely $[-1]$; but a principally polarized abelian variety with full level-N structure has no nontrivial automorphisms (provided that $N \geq 3$).

There are a number of ways to add structure to the set of curves of genus g. One method that works well is to add level structure to the Jacobian of the curve. A second method that is sometimes used is to specify the Weierstrass points or higher-order Weierstrass points, although this works well only in characteristic 0.

It is a deep fact that a coarse moduli space exists for curves of genus g or principally polarized abelian varieties of dimension g. These moduli

spaces are denoted by $\mathcal{M}_g$ and $\mathcal{A}_g$, respectively. It is not hard to see that if elements of M admit nontrivial automorphisms, then M cannot have a fine moduli space, hence the need to add additional structure. Thus there do not exist fine moduli spaces for curves of genus g or for principally polarized abelian varieties of dimension g, but there do exist fine moduli spaces for these classes of varieties with full level-N structure ($N \geq 3$). These fine moduli spaces are denoted by $\mathcal{M}_{g,N}$ and $\mathcal{A}_{g,N}$, respectively.

One can further show that $\mathcal{M}_{g,N}$ and $\mathcal{A}_{g,N}$ are quasi-projective, and that they have natural compactifications $\bar{\mathcal{M}}_{g,N}$ and $\bar{\mathcal{A}}_{g,N}$ carrying natural ample line bundles $\lambda_{\mathcal{M}}$ and $\lambda_{\mathcal{A}}$, respectively. Thus another way to define the height of a curve C or an abelian variety A is to set

$$h(C) = h_{\lambda_{\mathcal{M}}}([C]) \qquad \text{and} \qquad h(A) = h_{\lambda_{\mathcal{A}}}([A])$$

using the Weil heights associated to the line bundles $\lambda_{\mathcal{M}}$ and $\lambda_{\mathcal{A}}$. Here $[C]$ denotes the point of $\mathcal{M}_{g,N}$ corresponding to the isomorphism class of C and some choice of level-N structure on C, and similarly for $[A]$. We refer the reader to Mumford–Fogarty [1] for more precise statements, constructions, and properties of moduli spaces.

Parshin found an intrinsic way to define the height of an abelian variety over function fields, and this was extended to number fields by Faltings [1] and used in his proof of Mordell's conjecture. Let A be an abelian variety of dimension g defined over a number field k, let $\mathcal{A} \to \operatorname{Spec}(R_k)$ be the Néron model of A, let $\epsilon : \operatorname{Spec}(R_k) \to \mathcal{A}$ be the zero section, and let $\Omega = \Omega^g_{\mathcal{A}/\operatorname{Spec}(R_k)}$ be the bundle of relative g-differential forms on the scheme $\mathcal{A}$. The line bundle (invertible sheaf) $\omega_A := \epsilon^*(\Omega)$ on $\operatorname{Spec}(R_k)$ can be metrized in a natural way as follows.

Note that when restricted to $A(\mathbb{C})$, the bundle Ω of g-forms is equipped with the norm

$$\|\alpha\|^2 = \frac{1}{(2\pi)^g} \int_{A(\mathbb{C}} |\alpha \wedge \bar{\alpha}|.$$

The pullback of this norm via ε induces the desired metric on Ω_A. We thus have a metrized line bundle on $\operatorname{Spec}(R_k)$, and we can use the Arakelov degree to define the *Faltings height of A* to be

$$h_{\mathrm{Falt}}(A/k) := \frac{1}{[k:\mathbb{Q}]} \deg_{\mathrm{Ar}}(\omega_A, \|\cdot\|).$$

Let k' be a finite extension of k. If k'/k is unramified or if A/k is semistable, then $h_{\mathrm{Falt}}(A/k') = h_{\mathrm{Falt}}(A/k)$. In general, one always has the inequality $h_{\mathrm{Falt}}(A/k') \leq h_{\mathrm{Falt}}(A/k)$. These heights have a number of useful properties, for example,

$$h_{\mathrm{Falt}}(A \times B/k) = h_{\mathrm{Falt}}(A/k) + h_{\mathrm{Falt}}(B/k)$$

and

$$h_{\mathrm{Falt}}(\hat{A}/k) = h_{\mathrm{Falt}}(A/k).$$

(Here $\hat{A}$ denotes the dual of A.) There is also a nice formula describing how the Faltings height changes under isogeny. This formula plays a crucial role in Faltings' proof of the Mordell conjecture (see Section F.3.2 below). Another important result is a comparison theorem between the Faltings height and a suitably chosen Weil height $h_{\lambda_{\mathcal{A}}}$ on the moduli space $\mathcal{A}_g$ of abelian varieties. If A has everywhere semistable reduction, then

$$\left|h_{\text{Falt}}(A) - h_{\lambda_{\mathcal{A}}}([A])\right| \ll \log h_{\lambda_{\mathcal{A}}}([A]).$$

In particular, there are only a finite number of (principally polarized) abelian varieties with bounded height.

Yet another way to define the canonical height of an abelian variety is to use the canonical embedding defined by Mumford [7],

$$i_m = \phi_{\mathcal{L}^{\otimes m}} : A \hookrightarrow \mathbb{P}^{m^g - 1}.$$

This exhibits A as a subvariety of projective space, and one can then define (for a fixed even $m \geq 4$)

$$h_{\text{Theta}}(A) = h(i_m(A)).$$

Note that the right-hand side is the height (via Chow coordinates) of the subvariety $i_m(A)$ of $\mathbb{P}^{m^g-1}$. This height can also be compared to the other heights described above; see, for example, Bost–David [1].

Example F.2.2.1. Let E/k be an elliptic curve defined over a number field, let $\Delta_{E/k}$ be the minimal discriminant ideal of E/k, and let $j(E)$ be the j-invariant of E. (See Silverman [1] for basic definitions and formulas.) Then the Faltings height of E/k satisfies

$$h_{\text{Falt}}(E) \gg\ll \max\left\{h(j(E)), \log \mathrm{N}_{/k/\mathbb{Q}}\Delta_{E/k}\right\}.$$

In particular, if E/k is semistable, then

$$h_{\text{Falt}}(E) = \frac{1}{12}h(j(E)) + (\text{logarithmic error term}).$$

See Exercise F.5 and Silverman [9] for further information about the Faltings height of an elliptic curve.

Most of the above heights are intrinsically defined only up to bounded or constant functions. For example, the height of a projective variety $X \subset \mathbb{P}^n$ via Chow coordinates is defined only up to $O(\deg(X))$. Thus the height of a variety will depend on the choice of coordinates or models or metrics.

For abelian varieties, we know that the the group structure allows us to pick out a particularly good height for points, the canonical height. In a similar way, it is possible to define a theory of canonical heights for higher-dimensional subvarieties of abelian varieties. This is due to Philippon; see also Zhang [1] for a construction in the framework of Arakelov theory.

Lemma F.2.2.2. (Philippon [1]) *Let D be an ample divisor on an abelian variety A and let α be an endomorphism of A such that $\alpha^*D \sim q(\alpha)D$. For any subvariety X of A, let $G_X \subset A$ be the stabilizer of X. There is a constant $C = C(\alpha, D)$ such that for all subvarieties $X \subset A$,*

$$\left| h_D(\alpha X) - \frac{q(\alpha)^{\dim(X)+1}}{|\ker(\alpha) \cap G_X|} h_D(X) \right| \leq C(\alpha, D) \deg \alpha(X).$$

This lemma is formulated in Philippon [1] for a symmetric divisor and $\alpha = [n]$ (since then $[n]^* \sim n^2 D$), but the proof goes through verbatim in the more general case (see the remarks in David–Hindry [1]). It allows the following definition.

Definition. (Canonical height of a subvariety of an abelian variety)
Let D be a symmetric divisor on an abelian variety A and let X be a subvariety of A. The *canonical height of X with respect to D* is

$$\begin{aligned} \hat{h}_D(X) &= \lim_{n\to\infty} \frac{|\ker[n] \cap G_X|}{n^{2(\dim(X)+1)}} h_D([n](X)) \\ &= \lim_{n\to\infty} \frac{\deg(X)}{n^2 \deg([n](X))} h_D([n](X)). \end{aligned}$$

It is possible to define canonical heights with respect to any divisor, but this is a bit more involved, see the remark at the end of this section.

Theorem F.2.2.3. (Properties of canonical heights of subvarieties of abelian varieties)
The canonical height $\hat{h}_D$ depends only on the divisor class of D in Pic(A). *It further satisfies:*

(i) $\hat{h}_{mD}(X) = m^{\dim X+1} \hat{h}_D(X)$.
(ii) *If $t \in A$ is a torsion point, then $\hat{h}(t + X) = \hat{h}(X)$.*
(iii) *Suppose that $\alpha \in$ End(A) and that $\alpha^*(D) \sim q(\alpha)D$. Then*

$$\hat{h}_D(\alpha X) = \frac{q(\alpha)^{\dim(X)+1}}{|\ker \alpha \cap G_X|} \hat{h}_D(X)$$

and

$$\hat{h}_D(\alpha^{-1} X) = q(\alpha)^{\operatorname{codim}(X)-1} \hat{h}_D(X).$$

(iv) *$\hat{h}_D(X) = 0$ if and only if X is a torsion subvariety (i.e., if and only if $X = a + B$ with $a \in A_{\mathrm{tor}}$ and B an abelian subvariety of A).*

With the noteworthy exception of (iv), all of the properties in (F.2.2.3) are rather formal once the construction of $\hat{h}_D$ is achieved. Property (iv) lies much deeper, being in fact equivalent to the generalized Bogomolov conjecture stated in Section F.2 and proved by David–Philippon [1]. They even prove a lower bound of the form $\hat{h}(X) \geq c\,(\deg(X))^{-\kappa}$ with $c = c(A, k)$

and $\kappa = \kappa(g)$. The case where $A = E^n$ is a power of an elliptic curve was settled earlier by Philippon [1, Part III].

No simple formula connects $\hat{h}_{D+D'}(X)$ to $\hat{h}_D(X)$ and $\hat{h}_{D'}(X)$. Instead, for subvarieties of dimension r, there is an $(r+1)$-multilinear map that associates to each $(r+1)$-tuple $(D_0, \ldots, D_r)$ of symmetric divisors a canonical height $\hat{h}_{(D_0,\ldots,D_r)}(X)$. This height is linear in each D_i, and the canonical height defined above is equal to

$$\hat{h}_D(X) = \hat{h}_{(D,\ldots,D)}(X).$$

It is even possible to extend this to the case that s of the D_i's are symmetric and $r+1-s$ are antisymmetric. Then, by multilinearity, we obtain a map

$$\hat{h} : \operatorname{Pic}(A)^{r+1} \longrightarrow \left\{ \begin{matrix} \text{real-valued functions on the space of} \\ \text{subvarieties } X/\bar{\mathbb{Q}} \subset A \text{ of dimension } r \end{matrix} \right\}.$$

Finally, $\hat{h}_D$ can be defined for any divisor class by setting $\hat{h}_D = \hat{h}_{(D,\ldots,D)}$. See Gubler [1] for an Arakelov-style construction and de Diego's thesis (quoted in de Diego [1]) for a construction via Chow coordinates.

F.3. Height Bounds and Height Conjectures

A fruitful circle of ideas has been to try to link deep conjectures and results of Diophantine geometry to "elementary" statements such as the *abc* conjecture and Szpiro's conjecture. In this section we will discuss some of these conjectures and briefly indicate some of their consequences.

Definition. Let $n \neq 0$ be an integer. The *radical of n* is the product

$$\operatorname{rad}(n) = \prod_{p \mid n} p$$

of the primes dividing n. More generally, if k is a number field and $\alpha \in R_k$ an integer of k, then $\operatorname{rad}(\alpha)$, the *radical of* α, is the product over the prime ideals dividing the ideal (α).

The abc-Conjecture F.3.1. (Masser–Oesterlé) *For all $\varepsilon > 0$ there exists a constant $C_\varepsilon > 0$ such that if $a, b, c \in \mathbb{Z}$ are coprime integers satisfying $a + b + c = 0$, then*

$$\max\{|a|, |b|, |c|\} \leq C_\varepsilon \big(\operatorname{rad}(abc)\big)^{1+\varepsilon}.$$

(See the survey by Oesterlé [1] for a more complete discussion.)

This seemingly elementary conjecture has a tremendous number of far-reaching consequences. For example, we will sketch below Elkies' proof

that *abc* implies Faltings' theorem (Mordell's conjecture). Another interesting consequence is (asymptotic) Fermat's last theorem. For suppose that $x^p + y^p + z^p = 0$ for nonzero coprime integers x, y, z. Without loss of generality, we may assume that $|x| \le |y| \le |z|$. Then the *abc* conjecture implies that

$$\begin{aligned} |z|^p &= \max\{|x^p|, |y^p|, |z^p|\} \\ &\le C_\varepsilon \big(\operatorname{rad}(x^p y^p z^p)\big)^{1+\varepsilon} \le C_\varepsilon |xyz|^{1+\varepsilon} \le C_\varepsilon |z|^{3(1+\varepsilon)}. \end{aligned}$$

Since necessarily $|z| \ge 2$, we find that $p - 3(1+\varepsilon) \le \log_2(C_\varepsilon)$, and hence conclude that Fermat's equation has no nontrivial solutions if p is sufficiently large.

An earlier conjecture of Szpiro is closely related to the *abc* conjecture. In order to state Szpiro's conjecture, we recall a few definitions from the theory of elliptic curves. We will just give the definitions over $\mathbb{Q}$, and we refer the reader to Silverman [1, III §1, VII §§1–2, VIII §8] and Silverman [2, IV §10] for the generalization to number fields. An elliptic curve over $\mathbb{Q}$ has a Weierstrass equation

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$$

with $a_1, \dots, a_6 \in \mathbb{Z}$. The discriminant of this equation is a certain complicated polynomial in the a_i's (see Silverman [1, §3.1]), and a minimal Weierstrass equation for $E/\mathbb{Q}$ is one for which the the absolute value of this discriminant is minimized. The minimal discriminant Δ_E is the discriminant of a minimal Weierstrass equation for $E/\mathbb{Q}$. For primes $p|\Delta_E$, the reduction $\tilde{E}/\mathbb{F}_p$ of the Weierstrass equation modulo p will be singular. We say that E is semistable (or multiplicative) at p if the singularity is a node, and we say that E is unstable (or additive) at p if the singularity is a cusp. The conductor $\mathfrak{F}_E$ of $E/\mathbb{Q}$ is equal to

$$\mathfrak{F}_E = \prod_{p|\Delta_E} p^{\delta_p(E)}, \quad \text{where } \delta_p(E) = \begin{cases} 1 & \text{if } E \text{ is semistable at } p, \\ 2 & \text{if } E \text{ is unstable at } p \text{ and } p \ge 5. \end{cases}$$

If E is unstable at $p = 2$ or 3, the definition of $\delta_p(E)$ is more complicated, but in any case we always have $\delta_2(E) \le 8$ and $\delta_3(E) \le 5$.

For an elliptic curve E defined over a number field k, the minimal discriminant $\Delta_{E/k}$ and the conductor $\mathfrak{F}_{E/k}$ are integral ideals of k. The upper bounds for the exponents $\delta_\mathfrak{p}(E)$ for primes $\mathfrak{p}$ dividing 2 and 3 depend on the extent to which 2 and 3 are ramified in k. See Silverman [1, 2] cited above for the relevant definitions, and Lockhart–Rosen–Silverman [1] and Brumer–Kramer [1] for conductor bounds at small primes.

With these preliminaries, we can now state Szpiro's conjecture, and a related conjecture of Frey.

Conjecture F.3.2. *Let k be a number field. For all $\varepsilon > 0$, there is a constant $C = C(k, \varepsilon)$ such that for all elliptic curves E/k:*
(a) (Szpiro [3]) $\log|\Delta_{E/k}| \le (6+\varepsilon)\log \mathfrak{F}_{E,k} + C$.
(b) (Frey [1,2]) $h_k(j_E) \le (6+\varepsilon)\log \mathfrak{F}_{E/k} + C$.

It is not hard to show that Szpiro's conjecture, Frey's conjecture, and the *abc* conjecture are all more or less equivalent (up to some adjustment of the constants). See Exercise F.4 for some specific equivalences. We also recall that if E is semistable, then the Faltings height $h_{\rm Falt}(E)$ is equal (up to a logarithmic term) to $\frac{1}{12}h(j_E)$, so the conjecture can be reformulated using $h_{\rm Falt}$. Finally, if E is defined over $\mathbb{Q}$ and if we let c_4 and c_6 be the usual integers associated to a minimal Weierstrass equation for E, then a combined version of Conjecture F.3.2(a) and Conjecture F.3.2(b) can be stated using a naive height,

$$h_{\rm naive}(E) := \frac{1}{12}\log\max(c_4^4, c_6^2) \le (6+\varepsilon)\log \mathfrak{F}_{E/\mathbb{Q}} + C.$$

Remark F.3.3. Szpiro has suggested a generalization of Conjecture F.3.2 to abelian varieties. The conductor of an abelian variety A/k of dimension g is an ideal of the form $\mathfrak{F}_{A/k} = \prod \mathfrak{p}^{\delta_\mathfrak{p}}$, where the exponent $\delta_\mathfrak{p}$ is nonzero if and only if A has bad reduction at $\mathfrak{p}$. More precisely, it satisfies $0 \le \delta_\mathfrak{p} \le g$ if A has good or semistable reduction at $\mathfrak{p}$, and $\delta_\mathfrak{p} \le 2g$ if $\mathfrak{p}$ has characteristic greater than $2g+1$. More complicated bounds are known for primes of small characteristic. (See Brumer–Kramer [1] and Lockhart–Rosen–Silverman [1] for definitions and details.) The generalized Szpiro conjecture then asserts that there are constants c_1 and c_2, depending only on k and g, such that for all abelian varieties A/k of dimension g,

$$h_{\rm Falt}(A/k) \overset{?}{\le} c_1 \log \mathrm{N}_{k/\mathbb{Q}}\left(\mathfrak{F}_{A/k}\right) + c_2.$$

Notice that this gives a bound for $h_{\rm Falt}(A/k)$ solely in terms of $\dim(A)$ and the places of bad reduction. As we will see, an effective proof of this conjecture would provide an effective proof of Faltings' theorem (i.e., an effective bound for the height of rational of $C(k)$ on curves of genus$(C) \ge 2$).

Another conjecture of Lang, which seems at first glance to be unrelated to the above conjectures, postulates a uniform lower bound for the canonical height of nontorsion points on elliptic curves. This was strengthened and generalized to abelian varieties by Silverman.

Conjecture F.3.4. *Fix a number field k.*
(a) (Lang [5]) *There is a constant $c = c(k) > 0$ such that for all elliptic curves E/k and all nontorsion points $P \in E(k)$,*

$$\hat{h}(P) \ge c \log \mathrm{N}_{k/\mathbb{Q}}\Delta_{E/k}.$$

(b) (Silverman [5]) *Let $g \geq 1$. There is a constant $c = c(k,g) > 0$ such that for all abelian varieties A/k of dimension g, all ample divisors $D \in \mathrm{Div}_k(A)$, and all points $P \in A(k)$ such that the set $\{nP \mid n \in \mathbb{Z}\}$ is Zariski dense in A,*

$$\hat{h}(P) \geq ch(A).$$

(Note that the constant c will also depend on which notion of height is used for $h(A)$.)

Lang's conjecture (F.3.4) is known to be true for certain classes of elliptic curves, for example those with integral j-invariant (Silverman [5]). More generally, let

$$\sigma(E/k) = \frac{\log \mathrm{N}_{k/\mathbb{Q}} \Delta_{E/k}}{\log \mathrm{N}_{k/\mathbb{Q}} \mathfrak{F}_{E/k}}$$

be the *Szpiro ratio of E/k*. Then Hindry and Silverman [1] have proven that there is a constant $c = c\big([k : \mathbb{Q}], \sigma(E/k)\big) > 0$ such that for all nontorsion points $P \in E(k)$,

$$\hat{h}(P) \geq c \max\{h(j_E), \log \mathrm{N}_{k/\mathbb{Q}} \Delta_{E/k}\}.$$

Thus, since Szpiro's conjecture (F.3.2(a)) implies that $\sigma(E/k)$ is bounded, we see that Szpiro's conjecture implies Lang's conjecture (F.3.4(a)). In the higher-dimensional case, David [1] has proven a version of Conjecture F.3.4(b) for those abelian varieties A/k whose height $h(A)$ is bounded by a multiple of a Siegel period. David's result applies to infinitely many abelian varieties in each dimension.

An intriguing approach to the *abc* conjecture is through the theory of modular curves. The modularity conjecture, proven by Wiles [1] for semistable elliptic curves and ultimately extended to all elliptic curves $E/\mathbb{Q}$ by Breuil–Conrad–Diamond–Taylor [1], says that every elliptic curve $E/\mathbb{Q}$ admits a finite covering by a modular curve $\Phi_E : X_0(N) \to E$. Further, the integer N is the conductor of E, and if $\eta_E = dx/(2y + a_1x + a_3)$ is the invariant differential on a minimal Weierstrass equation for $E/\mathbb{Q}$ and $f_E(z)$ is the normalized weight-2 cusp form attached to E, then there is a rational number $c_E \in \mathbb{Q}^*$ such that

$$\Phi_E^*(\eta_E) = c_E f_E(z)\,dz.$$

The constant c_E is not too significant; for example, if E is semistable and has good reduction at 2, then $|c_E| = 1$ (Abbès–Ullmo [1]).

Integrating $\eta_E \wedge \overline{\eta}_E$ on $E(\mathbb{C})$ and pulling back via Φ_E gives a formula involving the Peterson norm of the modular form f,

$$\|f_E\|^2 := \frac{i}{2} \int_{X_0(N_E)} f_E(z)\,dz \wedge \overline{f_E(z)\,dz}.$$

After some computation one obtains the formula

$$\frac{1}{2}\log\deg(\Phi_E) = h_{\text{Falt}}(E/\mathbb{Q}) + \log\|f_E\| + \log|c_E|.$$

(See, e.g., Silverman [1, §3].) It is easy to prove that $\|f_E\|$ is bounded below by an absolute constant, so an estimate for the degree of the modular parametrization as in the following conjecture would suffice to prove a weak version of the *abc* conjecture.

Conjecture F.3.5. (Modular parametrization conjecture) *There is an absolute constant d such that for all (modular) elliptic curves $E/\mathbb{Q}$, there is a finite covering $\Phi_E : X_0(N) \to E$ with N equal to the conductor of E and* $\deg(\Phi_E) \le N^d$.

It appears that the best possible exponent in Conjecture F.3.5 is $d > 2$, so in order to deduce the full *abc* conjecture (i.e., with the constant $1+\varepsilon$), one would need a lower bound for $\|f_E\|$ of the form $\|f_E\| \gg N^{1/2-\varepsilon}$. It is not clear whether such an estimate is true, but Mestre and Oesterlé (unpublished) have shown that if N is square-free (i.e., E semistable), then $\|f_E\| \gg N^{1/4}$.

The Function Field Setting

Given the undoubted depth of the *abc* conjecture (F.3.1) and Szpiro's conjecture (F.3.2) over number fields, it is surprising how it easy it is to prove them over function fields. The statements and proofs have been discovered and rediscovered by numerous mathematicians, and we will not try to unsort the history here. For an elementary proof of Szpiro's conjecture over function fields due to Kodaira, see, for example, Hindry–Silverman [1, Theorem 5.1]. Similarly, Lang's height lower bound conjecture (F.3.4) can be proven unconditionally for function fields (Hindry–Silverman [1]), although the proof is more difficult.

By way of contrast, Elkies [1] has shown that over number fields, the *abc* conjecture implies Faltings' theorem (Mordell's conjecture). Elkies' proof does not carry over to the function field setting, because he utilizes a uniformization theorem of Belyi that does not have an appropriate function field analogue. We will sketch Elkies' proof below (Section F.4.2).

We close this section with a very short geometric proof of the *abc* conjecture for function fields shown to us by Bill Fulton. Before giving the proof, we recall that the degree of a nonconstant rational function f on a curve C is the degree of the associated finite map $f : C \to \mathbb{P}^1$. This is also equal to the number of zeros of f, taken with multiplicity; more generally, for any $\gamma \in \mathbb{P}^1$, $\deg(f) = \sum_{P\in f^{-1}(\gamma)} e_P(f)$. We also note that $\deg(f)$ is equal to the height $h(f)$ for the usual set of normalized valuations on the function field of C, $h(f) = \deg(f) = \sum_{P\in C}\max\{0, \operatorname{ord}_P(f)\}$.

Theorem F.3.6. (*abc* conjecture for function fields) *Let k be an algebraically closed field and let C/k be a smooth projective curve of genus g.*

Let $a, b \in k(C)$ be nonconstant functions satisfying $a+b=1$, and let $S \subset C$ be a set of points that includes all zeros and poles of a and b. Then

$$\deg(a) \leq \#S + 2g - 2.$$

PROOF. Let $d = \deg(a) = h(a)$. We apply the Riemann–Hurwitz formula (Theorem A.4.2.5) to the finite map $a : C \to \mathbb{P}^1$. This yields

$$\begin{aligned}
2g - 2 &= d(0-2) + \sum_{P \in C} (e_P(a) - 1) \\
&\geq -2d + \sum_{P \in a^{-1}(0)} (e_P(a) - 1) + \sum_{P \in a^{-1}(1)} (e_P(a) - 1) \\
&\qquad + \sum_{P \in a^{-1}(\infty)} (e_P(a) - 1) \\
&= -2d + \left(d - \#a^{-1}(0)\right) + \left(d - \#a^{-1}(1)\right) + \left(d - \#a^{-1}(\infty)\right) \\
&= d - \left(\#a^{-1}(0) + \#a^{-1}(1) + \#a^{-1}(\infty)\right) \\
&= d - \#\left(a^{-1}(0) \cup a^{-1}(1) \cup a^{-1}(\infty)\right).
\end{aligned}$$

We now observe that $a^{-1}(1) = b^{-1}(0)$, so

$$a^{-1}(0) \cup a^{-1}(1) \cup a^{-1}(\infty) = a^{-1}(0) \cup b^{-1}(0) \cup a^{-1}(\infty) \subset S.$$

Hence

$$2g - 2 \leq d - \#S,$$

which is the desired inequality. □

F.4. The Search for Effectivity

Here is a brief list of the main results of Diophantine geometry that we have proven in this book:

- **Mordell–Weil Theorem**
  The group of rational points on an abelian variety is finitely generated.
- **Roth's Theorem**
  There are only finitely many rational numbers $\alpha \in K$ that approximate a given irrational number β to within $H_K(\alpha)^{-2+\epsilon}$.
- **Siegel's Theorem**
  A curve of genus at least 1 has only finitely many S-integral points.
- **Faltings' Theorem**
  A curve of genus at least 2 has only finitely many rational points.

All of these statements are purely *qualitative*; that is, they merely assert that certain sets (of generators, numbers, or points) are finite. A major challenge is to make these theorems *effective*,(*) which means to give an effective procedure for computing all of the elements in the finite set. This generally means giving an effective upper bound for the heights of the elements in the set, since one knows that there are only a finite number of points of bounded height, and in principle it is then possible to list all of them and check which ones are actually in the set. None of these theorems has been proven effectively, although effective versions of Siegel's theorem are known for many classes of curves (e.g., for all elliptic curves) via techniques from transcendence theory and linear forms in logarithms. We also mention the related question of proving *quantitative* results, which means giving an explicit upper bound for the number of elements in the finite set. Quantitative versions of all of the above theorems are known.

F.4.1. Effective Computation of the Mordell–Weil Group $A(k)$

Let A/k be an abelian variety of dimension g over a number field k. The Mordell–Weil group of A/k is a finitely generated group (C.0.1),

$$A(k) \cong A(k)_{\text{tors}} \oplus \mathbb{Z}^{\operatorname{rank} A(k)}.$$

As we have seen in Theorem C.1.9, there is an effective upper bound for the rank in terms of k, g, and the places of bad reduction of A, but there is no effective procedure known for computing the rank exactly or for finding a set of generators for $A(k)$.

The torsion subgroup of $A(k)$ is much easier to deal with, and it is easy to give an effective algorithm to compute it. For example, if $E/\mathbb{Q}$ is an elliptic curve given by a Weierstrass equation

$$y^2 = x^3 + Ax + B$$

with integral coefficients, then the Lutz–Nagell theorem says that every torsion point $P = (x, y) \in E(\mathbb{Q})_{\text{tors}}$ has $x, y \in \mathbb{Z}$, and further, either $y = 0$ or else y^2 divides $4A^3 + 27B^2$. (See, e.g., Silverman [1, VIII.7.2].)

It is more difficult to give uniform bounds for the torsion subgroup, as in the following result.

Theorem F.4.1.1. *Let E be an elliptic curve defined over a number field k. We write C_m for a cyclic group of order m.*

(*) Also bear in mind that Matyasevic's negative solution to Hilbert's tenth problem says that not all Diophantine problems can be solved effectively. See Matyasevic [1] and Davis–Matyasevic–Putnam–Robinson [1].

(i) (Mazur [1]) *Let* $k = \mathbb{Q}$. *Then* $E(\mathbb{Q})_{\text{tors}}$ *is isomorphic to* C_m *with* $1 \le m \le 10$ *or* $m = 12$, *or to* $C_2 \times C_{2m}$ *with* $1 \le m \le 4$.

(ii) (Kamienny [1], see also Kenku-Momose [1]) *Suppose* $[k : \mathbb{Q}] = 2$. *Then* $E(\mathbb{Q})_{\text{tors}}$ *is isomorphic to* C_m *with* $1 \le m \le 16$ *or* $m = 18$, *or to* $C_2 \times C_{2m}$ *with* $1 \le m \le 6$, *or to one of* $C_3 \times C_3$, $C_4 \times C_4$, *or* $C_3 \times C_6$.

(iii) (Merel [1]) *In general, for every* $d \ge 1$ *there is a constant* C_d *such that for every number field* $k/\mathbb{Q}$ *with* $[k : \mathbb{Q}] \le d$ *and every elliptic curve* E/k, *we have* $\#E(k)_{\text{tors}} \le C_d$. *In particular, for each* d, *there are only a finite number of possible group structures for* $E(k)_{\text{tors}}$.

The proofs of all parts of Theorem F.4.1.1 are modular. For integers $m|M$, one studies the modular curve $X_1(m, M)$ whose (noncuspidal) points, denoted by $Y_1(m, M)$, classify isomorphism classes of pairs (E, ϕ), where E is an elliptic curve and ϕ is a map $\phi : C_m \times C_M \hookrightarrow E$ specifying a subgroup of E of type (m, M). The curve $X_1(m, M)$ is defined over $\mathbb{Q}(\zeta_m)$. When $m = 1$, it is customary to write $X_1(M)$ for $X_1(1, M)$; this curve is defined over $\mathbb{Q}$.

An elliptic curve E/k with a pair of independent k-rational points (P, Q) of orders m and M, respectively, corresponds to a point in the set $Y_1(m, M)(k)$. Thus Theorem F.4.1.1 describes those m and M for which $Y_1(m, M)(k) = \emptyset$. Note that the exceptional values in Mazur's theorem (F.4.1.1(a)) correspond exactly to those $X_1(m, M)$'s that are isomorphic to $\mathbb{P}^1$ over $\mathbb{Q}$. All of the exceptions in Kamienny's theorem (F.4.1.1(b)) have a similar geometric interpretation. The curves $X_1(3, 3)$ and $X_1(3, 6)$ are isomorphic to $\mathbb{P}^1$ over $\mathbb{Q}(\zeta_3)$, the curve $X_1(4, 4)$ is isomorphic to $\mathbb{P}^1$ over $\mathbb{Q}(i)$; the curves $X_1(11)$, $X_1(14)$, $X_1(15)$ and $X_1(2, 10)$ are of genus 1 and hence have infinitely many points over an infinity of quadratic fields; the curves $X_1(13)$, $X_1(16)$, and $X_1(18)$ are of genus 2, hence hyperelliptic and have infinitely many quadratic points.

A natural question raised by Theorem F.4.1.1 is whether an analogous result might hold for abelian varieties of higher dimension. There is little evidence today to suggest the correct answer, so we simply raise it as a question. For some results, examples, and discussion, see Silverberg's survey (Silverberg [1]) and the papers of Flynn [1] and Leprevost [1, 2].

Question F.4.1.2. *Let* $g \ge 1$ *and let* k *be a number field. Does there exist a constant* $C_{k,g}$ *such that for all abelian varieties* A/k *of dimension* g, *we have* $\#A(k)_{\text{tors}} \le C_{k,g}$? *If this is true, is it further possible to choose the constant depending only on* g *and the degree* $[k : \mathbb{Q}]$?

The rank of the Mordell–Weil group is much more mysterious and much more difficult to compute. Indeed, no one has yet devised an effective algorithm for computing the rank, or more generally for computing a set of generators. We will describe an algorithm of Manin [4] whose validity depends on a number of unproven conjectures of an analytic nature which we now discuss.

Recall that if we fix an ample symmetric divisor D on A, the canonical height $\hat{h}_D$ extends to a positive definite quadratic form on $A(k)\otimes\mathbb{R}$. (See Theorem B.5.3.) The associated bilinear form $\langle\cdot,\cdot\rangle_D$ is used to define the regulator

$$\operatorname{Reg}_D(A/k)=\det\bigl(\langle P_i,P_j\rangle_D\bigr)_{1\le i,j\le r},$$

where $P_1,\dots,P_r$ is a basis for $A(k)/A(k)_{\text{tors}}$. Thus the inner product $\langle\cdot,\cdot\rangle_D$ gives $A(k)\otimes\mathbb{R}\cong\mathbb{R}^r$ the structure of a Euclidean space, the image of $A(k)$ in this $\mathbb{R}^r$ is a lattice, and $\operatorname{Reg}_D(A/k)$ is the volume of a fundamental domain for the lattice.

Remark F.4.1.3. Rather than choosing a particular divisor D, we can define a canonical regulator of A/k as follows. Let $\hat{A}$ be the dual abelian variety to A, and choose a basis $P_1',\dots,P_r'$ for $\hat{A}(k)$. (Note that A and $\hat{A}$ are isogenous, so they have the same rank.) Let $\mathcal{P}$ be a Poincaré divisor on $A\times\hat{A}$ (see Section A.7.3), and let

$$\langle P,P'\rangle_{\mathcal{P}}=\hat{h}_{\mathcal{P}}(P,P')\qquad\text{for } (P,P')\in A\times\hat{A}$$

be the canonical height on $A\times\hat{A}$ with respect to $\mathcal{P}$. Then the *canonical regulator of A/k* is

$$\operatorname{Reg}(A/k):=\left|\det\bigl(\langle P_i,P_j'\rangle\bigr)_{1\le i,j\le r}\right|.$$

Since $\hat{h}_D(P)=\langle P,\phi_D(P)\rangle_{\mathcal{P}}$, and since $\phi_D:A\to\hat{A}$ is an isogeny if D is ample, it is easy to show that

$$\operatorname{Reg}_D(A/k)=\bigl[\hat{A}(k):\phi_D(A(k))\bigr]\operatorname{Reg}(A/k).$$

In particular, if D is a principal polarization, then $\operatorname{Reg}_D(A/k)=\operatorname{Reg}(A/k)$.

As indicated above, we have a lattice $A(k)/A(k)_{\text{tors}}$ sitting inside a Euclidean space $A(k)\otimes\mathbb{R}\cong\mathbb{R}^r$ with inner product $\langle\cdot,\cdot\rangle_D$. It is intuitively clear that we can bound the norm of some basis for the lattice if we know both an upper bound for the lattice's covolume (i.e., the volume of a fundamental domain) and a lower bound for the smallest nonzero vector in the lattice. The following result of Hermite makes this intuition precise.

Proposition F.4.1.4. (Hermite) *Let V be a real vector space of dimension r with Euclidean norm $\|\cdot\|$, let $L\subset V$ be a lattice, and let $\operatorname{Vol}(L)$ be the volume of a fundamental domain for L. Then there exists a basis $\mathbf{u}_1,\dots\mathbf{u}_r\in L$ for L satisfying*

$$\operatorname{Vol}(L)\le\|\mathbf{u}_1\|\cdot\|\mathbf{u}_2\|\cdots\|\mathbf{u}_r\|\le\left(\frac{4}{3}\right)^{r(r-1)/2}\operatorname{Vol}(L).\qquad(*)$$

Hence if the $\mathbf{u}_i$'s are ordered in increasing length, then

$$\|\mathbf{u}_i\| \le \left(\frac{4}{3}\right)^{r(r-1)/2(r-i+1)} \left(\frac{\operatorname{Vol}(L)}{\|\mathbf{u}_1\|}\right)^{1/(r-i+1)} \qquad \textit{for } 1 \le i \le r.$$

In particular, an upper bound for $\operatorname{Vol}(L)$ *and a (nonzero) lower bound for* $\|u\|$ *for nonzero vectors* $\mathbf{u} \in L$ *gives an upper bound for the length of a basis of* L.

PROOF. The left-hand inequality in $(*)$ is easy; it merely says that the volume of a parallelepiped is smaller than the product of its sides. For a proof of the right-hand inequality, see Lang [6, Chapter 5, Corollary 7.8]. □

In order to apply Hermite's result (F.4.1.4) to the Mordell–Weil lattice, we need a lower bound for $\hat{h}_D(P)$ and an upper bound for $\operatorname{Reg}_D(A/k)$. We discussed lower bounds for $\hat{h}_D(P)$ earlier; see (F.3.4). Here we will merely add the observation that in principle, one can compute an effective lower bound for $\hat{h}_D(P)$ by checking all points of bounded Weil height $h_D(P)$ and using an (effective) bound for the difference $\hat{h}_D - h_D$.

Currently, there is no proven algorithm to compute an upper bound for the regulator $\operatorname{Reg}_D(A/k)$, but a conjectural approach via zeta functions and analytic techniques was initiated by Manin [4]. We will describe this approach in some detail, but to avoid various technicalities we will treat only the case that A is defined over $\mathbb{Q}$.

Let $A/\mathbb{Q}$ be an abelian variety of dimension g. For each prime p, choose a prime $\ell \ne p$ and define the *Tate module of* A to be the inverse limit

$$T_\ell(A) = \varprojlim A[\ell^n]$$

relative to the maps $A[\ell^{n+1}] \xrightarrow{[\ell]} A[\ell^n]$. Let $V_\ell(A) = T_\ell(A) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$. Then $T_\ell(A) \cong \mathbb{Z}_\ell^{2g}$ and $V_\ell(A) \cong \mathbb{Q}_\ell^{2g}$. Further, there is a natural action of the Galois group $\operatorname{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ on these sets, which we denote by

$$\rho_\ell : \operatorname{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \longrightarrow \operatorname{Aut}_{\mathbb{Q}_\ell}\big(V_\ell(A)\big) \cong \operatorname{GL}(2g, \mathbb{Q}_\ell).$$

The map ρ_ℓ is the *ℓ-adic representation attached to $A/\mathbb{Q}$*.

For any prime $\mathfrak{p}$ over p, we let D_p and I_p denote respectively the decomposition and inertia groups of p. (Up to conjugation, D_p and I_p are independent of the choice of $\mathfrak{p}$; this ambiguity will not affect our discussion.) Thus $D_p = \{\sigma \mid \sigma(\mathfrak{p}) = \mathfrak{p}\}$, and I_p is the kernel of the reduction map $D_p \to \operatorname{Gal}(\bar{\mathbb{F}}_p/\mathbb{F}_p)$. This reduction map is surjective, and we let Frob_p denote an element of D_p that maps to the Frobenius $\alpha \mapsto \alpha^p$. It is well-defined up to an element of I_p (and up to conjugation).

Let $V_\ell(A)^{I_p}$ be the subspace of $V_\ell(A)$ fixed by every element of I_p. Then the action of Frob_p on $V_\ell(A)^{I_p}$ is well-defined up to conjugation, so its characteristic polynomial

$$Q_p(T) := \det\bigl(1 - \bigl(\rho_\ell(\mathrm{Frob}_p)\big|V_\ell(A)^{I_p}\bigr)T\bigr)$$

is well-defined. The *L-series of* $A/\mathbb{Q}$ is the Dirichlet series given by the product

$$L(A/\mathbb{Q}, s) = \sum_{n=1}^{\infty} a_n n^{-1} = \prod_p Q_p(p^{-1})^{-1}.$$

We remark that if A has good reduction at p, then the action of I_p on $V_\ell(A)$ is trivial. Thus for all but finitely many primes p we have $V_\ell(A)^{I_p} = V_\ell(A)$ and $\deg Q_p(T) = 2g$. This follows easily from the fact (C.1.4) that if A has good reduction at p, then the reduction map $A[\ell^n] \to \tilde{A}(\mathbb{F}_p)$ is injective. It is also known, but not at all obvious, that the polynomial $Q_p(T)$ is independent of the choice of auxiliary prime $\ell \neq p$.

If A has good reduction at p, then Weil's estimate says that $Q_p(T)$ factors over $\mathbb{C}$ as $Q_p(T) = \prod(1-\alpha_i T)$ with $|\alpha_i| = \sqrt{p}$. (For elliptic curves, this was originally proven by Hasse.) This immediately implies that the product defining $L(A/\mathbb{Q}, s)$ converges for $\mathrm{Re}(s) > \frac{3}{2}$. Conjecturally, much more is true.

Conjecture F.4.1.5.
(i) *The L-series $L(A/\mathbb{Q}, s)$ has an analytic continuation to all $s \in \mathbb{C}$.*
(ii) *There is an integer N, called the* conductor of A *and divisible by exactly the primes of bad reduction of A, such that if we set*

$$\Lambda(A/\mathbb{Q}, s) = N^{s/2}\bigl((2\pi)^{-s}\Gamma(s)\bigr)^g L(A/\mathbb{Q}, s),$$

then Λ satisfies the functional equation

$$\Lambda(A/\mathbb{Q}, 2-s) = \varepsilon\Lambda(A/\mathbb{Q}, s) \qquad \textit{for some } \varepsilon = \pm 1.$$

The number ε is called the sign of the functional equation.

Conjecture F.4.1.5 is known for abelian varieties with complex multiplication (Shimura–Taniyama [1]) and for quotients of the Jacobian $J_0(N)$ of $X_0(N)$ (Shimura [1]). In particular, Wiles [1] shows that every semistable elliptic curve $E/\mathbb{Q}$ is such a quotient; hence Conjecture F.4.1.5 is true for such curves. In a series of works by a number of mathematicians, Wiles's theorem was extended to ever larger collections of curves, culminating in the work of Breuil, Conrad, Diamond, Taylor [1] showing that it is true for every elliptic curve $E/\mathbb{Q}$, regardless of whether or not E has places of additive reduction.

Assuming the validity of (F.4.1.5), we may ask for the behavior of $L(A/\mathbb{Q}, s)$ near some special value of s, for example $s = 1$. The famous

conjecture of Birch and Swinnerton-Dyer gives an answer. We first set some notation. The *real period of* A is the integral $\Omega_A = \left|\int_{A(\mathbb{R})} \eta_A\right|$ of a Néron differential η_A (i.e., a differential normalized to have finite nonzero reduction on every fiber of the Néron model of A). For each prime p, let $A^0(\mathbb{Q}_p)$ denote the subgroup of $A(\mathbb{Q}_p)$ that reduces to the identity component of the Néron model of A, and let c_p be the index of $A^0(\mathbb{Q}_p)$ in $A(\mathbb{Q}_p)$. In particular, $c_p = 1$ for primes of good reduction, so $c_p = 1$ for all but finitely many primes. As usual, $\hat{A}$ is the dual abelian variety, and $Ш(A/\mathbb{Q})$ is the Tate–Shafarevich group, which we assume to be finite (see Section C.4).

Conjecture F.4.1.6. (Birch and Swinnerton-Dyer) *Assume that the* L*-series* $L(A/\mathbb{Q}, s)$ *admits an analytic continuation to* $\mathbb{C}$.

(i) *The order of vanishing of* $L(A/\mathbb{Q}, s)$ *at* $s = 1$ *is*

$$\operatorname*{ord}_{s=1} L(A/\mathbb{Q}, s) = \operatorname{rank} A(\mathbb{Q}).$$

(ii) *Let* $r = \operatorname{rank} A(\mathbb{Q})$. *Then the leading coefficient of the Taylor expansion of* $L(A/\mathbb{Q}, s)$ *around* $s = 1$ *is*

$$L^*(A, 1) := \lim_{s\to 1} \frac{L(A,s)}{(s-1)^r} = \Omega_A \Big(\prod_p c_p\Big) \frac{\#Ш(A/\mathbb{Q}) \cdot \operatorname{Reg}(A/\mathbb{Q})}{\#A(\mathbb{Q})_{\text{tors}} \cdot \#\hat{A}(\mathbb{Q})_{\text{tors}}}.$$

Birch and Swinnerton-Dyer originally formulated their conjecture for elliptic curves over $\mathbb{Q}$. It was then generalized by several people; the formulation given here comes essentially from Tate's Bourbaki seminar (Tate [3]). We also note that (F.4.1.5(ii)) and (F.4.1.6(i)) imply that the sign of the functional equation ε is equal to $(-1)^{\operatorname{rank} A(\mathbb{Q})}$, giving a comparatively easy (conjectural) analytic way to decide the parity of the rank.

Remark F.4.1.7. The conjecture of Birch and Swinnerton-Dyer may be viewed as an analogue of the classical formula for the residue of the Dedekind zeta function $\zeta_k(s)$ of a number field $k/\mathbb{Q}$. Let H_k be the class group of k, let Reg_k be the regulator of k, let D_k be the absolute value of the discriminant of k, let r_1 and r_2 be respectively the number of real and pairs of complex conjugate embeddings of k, and let μ_k be the group of roots of unity in k. Then

$$\lim_{s\to 1} (s-1)\zeta_k(s) = \frac{2^{r_1}(2\pi)^{r_2}}{\sqrt{D_k}} \cdot \frac{\#H_k \cdot \operatorname{Reg}_k}{\#\mu_k}.$$

(See, e.g., Lang [9], Theorem 5, Section VIII.2.) Comparing this formula to Conjecture F.4.1.6 suggests the following associations, although the reader is warned that the analogy is not perfect:

Number Field		Abelian Variety
R_k^*	$\longleftrightarrow$	$A(k)$
$\mu_k = (R_k^*)_{\text{tors}}$	$\longleftrightarrow$	$A(k)_{\text{tors}}$
Reg_k	$\longleftrightarrow$	$\operatorname{Reg}(A/k)$
H_k	$\longleftrightarrow$	$Ш(A/k)$

Remark F.4.1.8. If one assumes the analytic continuation and functional equation for the L-series of an elliptic curve (i.e., for $g = 1$), then one can prove that the series $\sum_{n=1}^{\infty} a_n n^{-s}$ actually converges for $\mathrm{Re}(s) > \frac{5}{6}$, and in particular in a neighborhood of $s = 1$. More generally, the series converges for $\mathrm{Re}(s) > (10g^2 - 6g + 1)/2g(4g - 1)$. See Chandrasekharan-Narasimhan [1] and Murty [1].

As pointed out by Manin [4], Conjectures F.4.1.5 and F.4.1.6 may be combined with analytic estimates to deduce information about the regulator $\mathrm{Reg}(A/\mathbb{Q})$.

Theorem F.4.1.9. *Let $A/\mathbb{Q}$ be an abelian variety of dimension g.*

(i) *Assume (F.4.1.5) that $L(A/\mathbb{Q}, s)$ has an analytic continuation and satisfies a functional equation, and further assume that its conductor N is at least 4. Let $r = \mathrm{ord}_{s=1} L(A/\mathbb{Q}, s)$. Then*

$$\left|L^*(A,1)\right| \le 2^r N^{\frac{1}{4}} (\log N)^{2g}.$$

(ii) *Assume further that the Birch–Swinnerton-Dyer Conjecture (F.4.1.6) is true. Then*

$$\begin{aligned}\mathrm{Reg}(A/\mathbb{Q}) &\le \#\text{Ш}(A/\mathbb{Q}) \cdot \mathrm{Reg}(A/\mathbb{Q}) \\ &\le 2^r N^{1/4} (\log N)^{2g} \Omega_A^{-1} \cdot \#A(\mathbb{Q})_{\mathrm{tors}} \cdot \#\hat{A}(\mathbb{Q})_{\mathrm{tors}}.\end{aligned}$$

All of the quantities on the right-hand side of this inequality can be effectively bounded from above, thereby giving an effective upper bound for $\mathrm{Reg}(A/\mathbb{Q})$, and hence via (F.4.1.4) for the height of a basis of $A(\mathbb{Q})$ (subject to the validity of Conjectures F.4.1.5 and F.4.1.6).

PROOF. The Hasse–Weil estimate says that the local factors used to define the L-series factor as $Q_p(T) = \prod(1 - \alpha_i T)$ with $|\alpha_i| = \sqrt{p}$. Writing $s = \sigma + i\tau$, this gives the upper bound

$$\left|Q_p(p^{-s})\right| \le (1 - p^{\frac{1}{2}-\sigma})^{-2g}.$$

If $\sigma > \frac{3}{2}$, we can multiply over p to obtain

$$\left|L(A/\mathbb{Q}, s)\right| \le \zeta\left(\sigma - \frac{1}{2}\right)^{2g}.$$

We also note that $\zeta(1+\varepsilon) \le 1 + 1/\varepsilon$ for all $\varepsilon > 0$.

Let $\Lambda(A/\mathbb{Q}, s) = N^{s/2}\left((2\pi)^{-s}\Gamma(s)\right)^g L(A/\mathbb{Q}, s)$ be the normalized L-series as in (F.4.1.5). Then the functional equation gives

$$\begin{aligned}\left|\Lambda\left(\frac{1}{2} - \varepsilon - i\tau\right)\right| &= \left|\Lambda\left(\frac{3}{2} + \varepsilon + i\tau\right)\right| \\ &\le \left(\Gamma\left(\frac{3}{2} + \varepsilon\right)(2\pi)^{-\frac{3}{2}-\varepsilon}\right)^g N^{\frac{3}{4}+\frac{\varepsilon}{2}} (1 + 1/\varepsilon)^{2g}.\end{aligned}$$

By the Phragmén–Lindelöf principle, the same estimate is valid in the vertical strip $\frac{1}{2}-\varepsilon \le \sigma \le \frac{3}{2}+\varepsilon$. Applying Cauchy's inequality (to the circle with center at 1 and radius $\frac{1}{2}+\varepsilon$ with $\varepsilon = 2/\log N$), we get after some calculation

$$L^*(A,1) = \frac{(2\pi)^g}{\sqrt{N}} \cdot \frac{\Lambda^{(r)}(1)}{r!} \le \frac{(2\pi)^g}{\sqrt{N}} \cdot \frac{\max|\Lambda(s)|}{\left(\frac{1}{2}+\varepsilon\right)^r} \le 2^r N^{\frac{1}{4}} \log N^{2g}.$$

□

Remark F.4.1.10. The assumption in (F.4.1.9) that the conductor N satisfies $N \ge 4$ is not needed. Fontaine [1] has proven that if $A/\mathbb{Q}$ is an abelian variety of dimension g, then its conductor satisfies $N > 10^g$. In particular, for elliptic curves the smallest possible conductor is $N = 11$, a fact first proven by Tate. There are three isomorphism classes of elliptic curves of conductor 11, two of which are the classical modular curves $X_0(11)$ and $X_1(11)$. All three are isogenous to one another. Note that if E has conductor 11, then $A = E^g$ has conductor 11^g, so the lower bound of 10^g is reasonably sharp.

Remark F.4.1.11. The Faltings height $H(A) := \exp\bigl(h_{\text{Falt}}(A)\bigr)$ is essentially a complex period (Exercise F.6), so it may be compared to the real period Ω_A. One finds that $\Omega_A^{-1} = H(A)\sqrt{\det \operatorname{Im}\tau}$, where τ is the point in the Siegel upper half-plane determined by the period matrix of A. (See, for example, the matrix lemma in Masser [1].) Hence

$$\det \operatorname{Im}\tau \le \|\operatorname{Im}\tau\|^g \ll \bigl(\log H(A)\bigr)^g \quad \text{and} \quad \Omega_A^{-1} \ll H(A)\bigl(\log H(A)\bigr)^{g/2}.$$

These calculations and Conjecture F.3.4 lead us to expect that the smallest set of generators $\{P_1,\dots,P_r\}$ for the free part of the Mordell–Weil group $A(\mathbb{Q})$ of a simple abelian variety will always satisfy

$$\log H(A) \ll \min_{1\le i\le r} \hat{h}(P_i) \le \max_{1\le i\le r} \hat{h}(P_i) \ll H(A)^{1+\varepsilon},$$

and further that both sides of this inequality are, in some cases, close to the truth. The wide range of values provides, if not an explanation, at least an illustration of the difficulty of computing the Mordell–Weil group.

Remark F.4.1.12. As we have seen (F.4.1.1), the torsion subgroup of an elliptic curve is subject to a universal bound. In stark contrast, there is the following folklore conjecture.

Conjecture F.4.1.13. *There exist elliptic curves $E/\mathbb{Q}$ whose Mordell–Weil rank* $\operatorname{rank} E(\mathbb{Q})$ *is arbitrarily large. More generally, for any $g \ge 1$ there exist geometrically simple abelian varieties $A/\mathbb{Q}$ with* $\operatorname{rank} A(\mathbb{Q})$ *arbitrarily large.*

As of this writing, the record for an elliptic curve is $\operatorname{rank} E(\mathbb{Q}) \ge 23$ (Martin and McMillen, June 1997).

It is not difficult to prove an effective unconditional upper bound of the form $\operatorname{rank} E(\mathbb{Q}) \ll \log N$, where N denotes the conductor of E. If $E(\mathbb{Q})_{\text{tors}}$ is not empty, the stronger bound

$$\operatorname{rank} E(\mathbb{Q}) \ll \frac{\log N}{\log \log N}$$

can be proven unconditionally in a similar manner. If $E(\mathbb{Q})_{\text{tors}}$ is empty, Mestre [1] has shown that the stronger estimate is valid if one assumes a number of standard conjectures, including (F.4.1.5), (F.4.1.6), and the generalized Riemann hypothesis for $L(E/\mathbb{Q}, s)$.

F.4.2. Effective Computation of Rational Points on Curves

Let C/k be a curve of genus $g \geq 2$ defined over a number field k. Faltings' theorem says that the set of rational points $C(k)$ is finite. An effective version of Faltings' theorem would provide an explicit upper bound $B = B(C, k)$ in terms of C and k such that

$$P \in C(k) \Longrightarrow h(P) \leq B.$$

At present, effective bounds of this sort are not known. In this section we will briefly discuss the following approaches to giving an effective proof of Faltings' theorem:

- **The Naive Approach**
  A restatement of the problem in elementary terms.
- **The Mordell–Weil Group Approach**
  Embed C in its Jacobian variety.
- **The Arakelov Theory Approach**
  Use Arakelov intersection theory on an arithmetic surface.
- **The Moduli Approach**
  Associate points in $C(k)$ to points on a moduli space.
- **The *abc* Approach**
  Use the *abc* conjecture and uniformization.
- **The Small Point Approach**
  Use hypothetical small points in $C(\bar{k})$.

The Naive Approach
An algebraic curve always has an affine plane model $P(X,Y) = 0$, with at worst simple nodes as singularities. One then searches for a constant $B = B(P, k)$ such that any solution to $P(x, y) = 0$ with $x, y \in k$ satisfies $h(1, x, y) \leq B$.

The Mordell–Weil Group Approach

The canonical height $\|x\|^2 := \frac{1}{2}\hat{h}_{\Theta+\Theta^-}(x)$ gives a Euclidean norm on $\operatorname{Jac}(C)(k)$. We can also map C canonically to its Jacobian via the map

$$j : C \longrightarrow \operatorname{Jac}(C), \qquad P \longmapsto \operatorname{Cl}\big((2g-2)P - K_C\big).$$

One then looks for a bound $B = B(C,k)$ such that any point $P \in C(k)$ satisfies $\|j(P)\| \le B$.

This is essentially the method used in the Bombieri–Vojta proof described in Part E. However, rather than producing a bound B that works for all points, the proof in Part E gives only a pair of effective bounds $N = N(C,k)$ and $B = B(C,k)$ such that there are at most N points in $C(k)$ satisfying $h(P) \ge B$. The reason for this lack of effectivity is rooted in the very nature of Diophantine approximation proofs, wherein one postulates the existence of a few large solutions and uses them to bound the size of all remaining solutions. This means that the proof provides no way of checking whether there are any large solutions at all. Notice that a similar remark applies to the proof of Roth's theorem in Part D.

The Arakelov Theory Approach

The curve C/k is associated to an arithmetic surface $\pi : \mathcal{C} \to \operatorname{Spec}(R_k)$ that comes equipped with a dualizing sheaf $\omega_{\mathcal{C}/R_k}$. (This is essentially the relative canonical sheaf of $\mathcal{C}$ over R_k.) Each point $P \in C(k)$ corresponds to a section $\sigma_P : \operatorname{Spec}(R_k) \to \mathcal{C}$ of π. Let $E_P = \sigma_P(R_k)$ denote the image, and define the Arakelov height of P to be

$$h_{\mathrm{Ar}}(P) = \frac{1}{[k:\mathbb{Q}]} E_P \cdot \omega_{\mathcal{C}/R_k} = \frac{1}{[k:\mathbb{Q}]} \deg_{\mathrm{Ar}} P^*\omega_{\mathcal{C}/R_k}.$$

One looks for a bound B such that $h_{\mathrm{Ar}}(P) \le B$ for all $P \in C(k)$, where B is given in terms of quantities such as the self-intersection $\omega_{\mathcal{C}/R_k} \cdot \omega_{\mathcal{C}/R_k}$ or $h_{\mathrm{Falt}}(\operatorname{Jac}(C))$ or the set of places of bad reduction. A proof along these lines would be aesthetically pleasing, due to the canonical nature of all of the quantities involved.

The Moduli Approach

This is the method used by Faltings [1] (see also Szpiro [1,2,3]) in his original proof of the Mordell conjecture. The idea is to associate to each point $P \in C(k)$ another variety (in this case, a curve) X_P and then prove that there are only finitely many such X_P's. The first step is the following construction of Kodaira and Parshin.

Proposition F.4.2.1. (Kodaira–Parshin construction) *Let C/k be a curve of genus $g \ge 2$ defined over a number field. There exists a smooth projective surface X with a fibration $\pi : X \to C$ whose fibers $X_P = \pi^{-1}\{P\}$ are smooth curves of genus g'. The fibration has the following two properties.*

(i) *The function* $P \mapsto h_{\text{Falt}}(\text{Jac}(X_P))$ *is a height function on* $C(\bar{\mathbb{Q}})$ *relative to an ample divisor.*

(ii) *There exists a finite set of places* S *of* k *such that* X_P *has good reduction outside of* S *for every point* $P \in C(k)$.

The Kodaira–Parshin construction can be made effective in the sense that the set S in (ii) can be explicitly described in terms of the places of bad reduction of C, and the height in (i) satisfies

$$\left|h_{\text{Falt}}(\text{Jac}(X_P)) - c_1 h_{\text{Ar}}(P)\right| \le c_2,$$

where c_1 is an explicit rational number and c_2 is effective (although difficult to compute). Hence all that we need to make Faltings' theorem effective is an effective bound for $h_{\text{Falt}}(\text{Jac}(X_P))$, and from (ii) it suffices to bound the height of abelian varieties of dimension g' having good reduction outside S. The fact that this set is finite was conjectured by Shafarevich and proven by Faltings [1].

Let $\mathcal{A}(g,k,S)$ denote the set of abelian varieties of dimension g defined over k and having good reduction outside S. Faltings proves that $\mathcal{A}(g,k,S)$ is finite in two steps:

(F1) *The set* $\mathcal{A}(g,k,S)$ *contains only a finite number of isogeny classes.*

(F2) *Fix an abelian variety* A/k. *Then there are only finitely many (isomorphism classes of) abelian varieties defined over* k *and isogenous to* A.

The proof of (**F2**) by Faltings uses delicate arguments involving finite group schemes and p-divisible groups. Faltings' proof has been refined by Raynaud to make it effective (see Raynaud's paper in Szpiro [2]).

To each isogeny $\alpha : A \to B$ of degree d defined over k there is an ideal δ_α in R_k such that

$$h(B) = h(A) + \frac{1}{2}\log d - \log|R_k/\delta_\alpha| \qquad \text{and} \qquad \delta_\alpha\delta_{\hat{\alpha}} = dR_k,$$

where $\hat{\alpha} : \hat{B} \to \hat{A}$ denotes the dual isogeny. This gives the inequality

$$\left|h(A) - h(B)\right| \le \frac{1}{2}\log d.$$

Masser–Wüstholz [2, 3] have shown by a totally different method based on transcendence theory techniques that if B is isogenous to A, then there exists an isogeny of degree $d \le \kappa h(A)^\lambda$, where κ and λ are explicitly given functions of $g = \dim A$ and $[k : \mathbb{Q}]$. This provides an effective proof of (**F2**) with additional uniformity.

Bost [2] has given a simpler method of comparing the heights of abelian varieties, thereby providing an easier proof of the finiteness of abelian varieties with bounded Faltings height. For example, if (A, L) is a principally polarized abelian variety, Bost describes a normalized Chow height

$h_{\text{norm}}(A, L)$ such that $h_{\text{norm}}(A, L) \le h_{\text{Falt}}(A) + \frac{g}{4}\log(2\pi)$. (See also Bost–David [1]).

Thus to make Faltings' theorem effective, it suffices to give an effective bound for the height of a representative abelian variety in each isogeny class of abelian varieties with good reduction outside of S. In other words, give an effective proof of (**F1**). Unfortunately, no such proof is known today. Faltings' proof of (**F1**) is beautiful, but indirect. He first proves Tate's conjecture that A and B are isogenous if and only if their Tate modules $T_\ell(A)$ and $T_\ell(B)$ are isomorphic as $\text{Gal}(\bar{k}/k)$-modules. He also proves that the $T_\ell(A)$ is semisimple. These results are obtained by an adaptation of methods of Tate and Zarhin and rely on the proof of (**F2**).

The ℓ-adic representation

$$\rho_{A,\ell} : \text{Gal}(\bar{k}/k) \longrightarrow \text{Aut}_{\mathbb{Q}_\ell}\bigl(T_\ell(A) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell\bigr) \cong \text{GL}(2g, \mathbb{Q}_\ell)$$

attached to A has a number of remarkable properties, including:

(i) For each place v of k, the characteristic polynomial of $\rho_{A,\ell}(\text{Frob}_v)$ is a polynomial with integer coefficients;

(ii) The roots of this characteristic polynomial (i.e., the eigenvalues of $\rho_{A,\ell}(\text{Frob}_v)$) all have absolute value $q_v^{1/2}$ (Weil's theorem).

It follows that the trace of $\rho_{A,\ell}(\text{Frob}_v)$ can take on only finitely many values. Faltings then completes the proof with a wonderful lemma saying that a semisimple representation $\rho : \text{Gal}(\bar{k}/k) \to \text{GL}(2g, \mathbb{Q}_\ell)$ that is unramified outside S is determined by a finite number of trace values $\rho(\text{Frob}_v)$. (See Faltings [1, Satz 5] or Szpiro [2, Théorème 1 on page 249].)

Thus the missing piece in making this proof of Faltings' theorem effective is an effective bound $B = B(g, k, S)$ with the property that each isogeny class in $\mathcal{A}(g, k, S)$ contains an abelian variety A/k satisfying $h(A) \le B$. Tate's conjecture, proven by Faltings [1], says that the isogeny class of A is characterized by its L-series $L(A/k, s)$. Further, if A has good reduction outside S, then the conductor of A is bounded by a constant depending only on g and the primes in S, which (at least conjecturally) provides further information about $L(A/k, s)$. It is thus tempting to ask whether it might be possible to relate the height $h(A)$ to the L-series $L(A/k, s)$ (or some variant) and thereby bound the height in terms of the conductor.

The *abc* Approach

The *abc* conjecture (F.3.1) is an elementary-sounding statement about primes dividing coprime integers (a, b, c) satisfying $a + b + c = 0$. Elkies [1] has explained how the *abc* conjecture and a uniformization theorem of Belyi [1] can be used to give a short proof of the finiteness of $C(k)$ for any curve C/k of genus $g \ge 2$. Further, an effective proof of the *abc* conjecture would yield an effective proof of Faltings' theorem. We will sketch the proof for $k = \mathbb{Q}$, but the arguments are valid for any number field.

We begin by translating the *abc* conjecture into a form more suitable for our purposes. For any rational number $x \in \mathbb{Q} \smallsetminus \{0,1\}$, let

$$N_0(x) = \prod_{\operatorname{ord}_p(x)>0} p, \qquad N_1(x) = \prod_{\operatorname{ord}_p(x-1)>0} p, \qquad N_\infty(x) = \prod_{\operatorname{ord}_p(x)<0} p,$$

and set

$$N(x) = N_0(x)N_1(x)N_\infty(x) \qquad \text{and} \qquad H(x) = H(1,x).$$

Then an alternative statement of the *abc* conjecture says that for every $\varepsilon > 0$ there exists a $C_\varepsilon > 0$ such that

$$N(x) \geq C_\varepsilon H(x)^{1-\varepsilon} \qquad \text{for all } x \in \mathbb{Q} \smallsetminus \{0,1\}.$$

(To see the equivalence, set $x = -a/b$.)

Next let $C/\mathbb{Q}$ be a curve of genus $g \geq 2$. Belyi's theorem [1] (see also Exercise A.4.7) says that there is a finite map $f : C \to \mathbb{P}^1$, say of degree d, that is ramified only above the three points $\{0, 1, \infty\}$. Letting $m := \#\big(f^{-1}\{0,1,\infty\}\big)$, the Riemann–Hurwitz formula (Theorem A.4.2.5) gives

$$\begin{aligned} 2g-2 &= d\cdot(-2) + \sum\big(e_P(f)-1\big) \\ &= -2d + \big(d - \#f^{-1}(0)\big) + \big(d - \#f^{-1}(1)\big) + \big(d - \#f^{-1}(\infty)\big) \\ &= d - m. \end{aligned}$$

Later we will choose $\varepsilon < (2g-2)/d$, which will guarantee that $m/d < 1-\varepsilon$.

Now we study points outside $f^{-1}\{0,1,\infty\}$. Let

$$D_0 = (f)_0 = \sum_{\operatorname{ord}_Q(f)>0} \operatorname{ord}_Q(f)(Q) \qquad \text{and} \qquad D_0' = \sum_{\operatorname{ord}_Q(f)>0} (Q).$$

In other words, D_0 is the divisor of zeros of f taken with multiplicities, and D_0' is the same without multiplicities. In particular, $d = \deg(D_0)$, and we will write $d_0' = \deg(D_0')$. The divisor $d_0'D_0 - dD_0'$ has degree 0, so is algebraically equivalent to 0 on C, and D_0 is ample, so Theorem B.5.9 and a little algebra using the height machine (B.3.2) gives us the height relation

$$h_{D_0'} = \frac{d_0'}{d} h_{D_0} + O\Big(\sqrt{h_{D_0}}\Big).$$

Let $P \in C(\mathbb{Q})$ with $f(P) \neq 0, \infty$. A prime p will occur in the numerator of the rational number $f(P)$ if and only if it contributes to the height

$H_{D_0'}(P)$, so we obtain an inequality $N_0(f(P)) \ll H_{D_0'}(P)$. Substituting this into the above height relation gives the estimate

$$\begin{aligned}\log N_0(f(P)) &\le \frac{d_0'}{d} h_{D_0}(P) + O\Big(\sqrt{h_{D_0}(P)}\Big)\\ &= \frac{d_0'}{d} h(f(P)) + O\Big(\sqrt{h(f(P))}\Big).\end{aligned}$$

Repeating this argument using the divisors $D_1 = (f-1)_0$ and $D_\infty = (1/f)_0 = (f)_\infty$ and the multiplicity-free degrees $d_1' = \#(f^{-1}\{1\})$ and $d_\infty' = \#(f^{-1}\{\infty\})$ gives analogous inequalities

$$\begin{aligned}\log N_1(f(P)) &\le \frac{d_1'}{d} h(f(P)) + O\Big(\sqrt{h(f(P))}\Big),\\ \log N_\infty(f(P)) &\le \frac{d_\infty'}{d} h(f(P)) + O\Big(\sqrt{h(f(P))}\Big).\end{aligned}$$

Adding the three inequalities and noting that $d_0' + d_1' + d_\infty' = m$ yields

$$\log N(f(P)) \le \frac{m}{d} h(f(P)) + O\Big(\sqrt{h(f(P))}\Big).$$

On the other hand, the *abc* conjecture tells us that for any $\varepsilon > 0$ there is a constant c_ε such that

$$\log N(f(P)) \ge (1-\varepsilon) h(f(P)) - c_\varepsilon.$$

Hence there is a constant c_ε', depending on c_ε and the above big-O constant), such that

$$\left(1 - \varepsilon - \frac{m}{d}\right) h(f(P)) \le c_\varepsilon'.$$

In particular, if we choose $\varepsilon < (2g-2)/d$, then $m/d < 1 - \varepsilon$, and we obtain a nontrivial upper bound for $h(f(P))$. Further, if the constant in the *abc* conjecture could be made effective, then this proof would give an effective bound for the height of points in $C(\mathbb{Q})$.

The Small Point Approach

Another conjectural approach to an effective proof of Faltings' theorem, more in the spirit of Arakelov geometry, was proposed by Szpiro [1, 2, 3]. It is based on the following conjecture that every curve has an algebraic point of "small" height.

Conjecture F.4.2.2. (Szpiro's small points conjecture) *Let $g \ge 2$, let k be a number field, and let S be a finite set of places of k. There exists a constant $B(k, g, S)$ with the following property.*

Let C/k be a curve of genus g with good reduction outside of S. Then C contains an algebraic point $P \in C(\bar{k})$ whose Arakelov self-intersection is bounded by

$$-\frac{1}{[k(P):\mathbb{Q}]}E_P^2 \leq B(k,g,S).$$

Here E_P denotes the horizontal divisor associated to P on a minimal model for C over $\operatorname{Spec} R_{k(P)}$.

Szpiro observes that Conjecture F.4.2.2 is true in the function field case, where one may take $B(k,g,S) = 2g - 2 + \#S + 1$. The fact that a statement such as (F.4.2.2) implies a bound for the height of rational points on curves follows from an analysis of the Kodaira–Parshin fibration (F.4.2.1) $\pi : X \to C$, which yields an inequality

$$-\frac{1}{[k(P):\mathbb{Q}]}E_P^2 \leq c_1 \inf_{Q \in X_P(\bar{k})} \left\{ -\frac{1}{[k(Q):\mathbb{Q}]}E_Q^2 \right\} + c_2$$

with absolute constants c_1 and c_2. For a further discussion on effectivity and the relationship with bounds for $\omega^2_{\mathcal{C}/R}$ and associated quantities, we refer the reader to Szpiro's original papers [1, 2, 3], Parshin [1], and Moret-Bailly [1].

Finally, our discussion of effectivity for rational points on curves would not be complete without the confession that even Siegel's theorem (D.9.1) on integral points is not yet effective. More precisely, let C be a smooth affine curve of genus g, let $\bar{C}$ be a smooth projective closure of C, and let $s = \#(\bar{C} \smallsetminus C)$ be the number of points "at infinity" on C. Siegel's theorem says that if $2g - 2 + s > 0$, then the set of S-integral points $C(R_{k,S})$ is finite. This theorem is effective when $g = 0$ (and $s \geq 3$) and when $g = 1$ (and $s \geq 1$) thanks to Baker's theorem on lower bounds for linear forms in logarithms. (See, e.g., Baker [1] or Serre [3].) There are many other families of curves for which Siegel's theorem can be made effective by similar techniques, including, for example, curves of the form $Y^m = P(X)$ and Thue curves $F(X,Y) = a$ for a homogeneous form F. Nevertheless, despite the fact that we know that there are only finitely many rational points on curves of genus $g \geq 2$, there is still no general effective proof for the finiteness of integral points.

F.4.3. Quantitative Bounds for Rational Points

In the last two sections we discussed effective bounds for generators of the Mordell–Weil group and for rational points on curves. A related, but easier, problem is to give an explicit upper bound for the number of elements in these finite sets, without necessarily bounding the height of their elements. Estimates of this nature are called *quantitative bounds.*

For example, we have already given the following quantitative version of the Mordell–Weil theorem (C.1.9):

$$\operatorname{rank} A(k) \le 2g \operatorname{rank} R_{k,S}^* = 2g(r_1 + r_2 + \#S - 1).$$

In this estimate, it is assumed that $A[m] \subset A(k)$ for some $m \ge 2$ and that $R_{k,S}$ is a principal ideal domain, but in general it is not hard to replace k and S by an explicit extension k'/k and an explicit expanded set S' having these properties. In this way one obtains an upper bound for $\operatorname{rank} A(k)$ solely in terms of g, k, and the primes of bad reduction of A.

Similarly, Roth's theorem, Siegel's theorem, and Faltings' theorem, each of which asserts that a certain set $\mathcal{S}$ is finite, can be given quantitative formulations of the following sort:

There are effective constants c_1 and c_2 such that
$\mathcal{S} = \{x \in \mathcal{S} \mid h(x) \le c_1\} \cup \mathcal{S}' \quad$ with $\#\mathcal{S}' \le c_2$.

The explicit form taken by the constants c_1 and c_2 is naturally of interest. To illustrate this type of result, we quote a quantitative version of Faltings' theorem due de Diego. The proof is an adaptation of that given in Part E. We also note that since it is possible to include all curves of genus g into a huge algebraic family, Theorem F.4.3.1 in principle gives an explicit upper bound for $\#C(k)$ for all curves C of genus $g \ge 2$.

Theorem F.4.3.1. (de Diego [1]) *Let $f : X \to T$ be a family of smooth curves $X_t := f^{-1}\{t\}$ of genus $g \ge 2$. Fix a height function h_T on the base T, and for each $t \in T$, let h_t be the height function on X_t associated to the pullback of the canonical height on* $\operatorname{Jac}(X_t)$ *with respect to $\Theta + \Theta^-$. Then there is an effectively computable constant c, depending on h and $f : X \to T$ (but not on the number field k), such that for every $t \in T(k)$,*

$$X_t(k) = \{x \in X_t(k) \mid h_t(x) \le c h_T(t)\} \cup \mathcal{E}_t$$

with an exceptional set $\mathcal{E}_t$ satisfying $\#\mathcal{E}_t \le 7^{2+\operatorname{rank}\operatorname{Jac}(X_t)(k)}$.

In Theorem F.4.3.1, the size of the exceptional set is bounded in terms of the rank of the Jacobian variety, but one might suspect that the exceptional set is not actually necessary. This would lead to the following precise (effective) form of Faltings' theorem.

Conjecture F.4.3.2. *In the situation of Theorem F.4.3.1, it is possible to choose the constant c in such a way that the exceptional set $\mathcal{E}_t$ is empty. In other words, there is a uniform upper bound $h_t(x) \le c h_T(t)$ that holds for all $t \in T(k)$ and all $x \in X_t(k)$.*

In the opposite direction, one might ask for some sort of uniform bound for the number of points of small height in $X_t(k)$. One approach to doing this is the following elementary remark (see Exercise B.10). Let A/k be an abelian variety, let $\hat{h}$ be the canonical height on A with respect to an ample symmetric divisor, let $r = \operatorname{rank} A(k)$, and let $\hat{h}_{\min}$ be the minimum value of $\hat{h}(P)$ for nontorsion points $P \in A(k)$. Then

$$\#\{x \in A(k) \mid \hat{h}(x) \le B\} \le \#A(k)_{\mathrm{tors}} \cdot \left(1 + \sqrt{B/\hat{h}_{\min}}\,\right)^r.$$

Thus one way to prove a good quantitative bound for a subset $\mathcal{S}$ of $A(k)$ is to find an upper bound B for the heights of elements in $\mathcal{S}$ that is proportional to the lower bound $\hat{h}_{\min}$ of the heights of nontorsion points in $A(k)$.

The following two conjectures illustrate ways in which one might hope to uniformly bound the number of integral points on elliptic curves and the number of rational points on curves of higher genus.

Conjecture F.4.3.3. *Let k be a number field.*

(i) (Lang [5]) *There is a constant $c = c(k)$ such that if E is a minimal affine model of an elliptic curve over R_k and S is a finite set of places of k, then*

$$\#E(R_{k,S}) \le c^{1+\operatorname{rank} E(k) + \#S}.$$

(ii) (Mazur) *There is a constant $c = c(k, g)$ such that if C is a curve of genus $g \ge 2$ defined over k, then*

$$\#C(k) \le c_2^{1+\operatorname{rank} \operatorname{Jac}(C)(k)}.$$

Silverman [3] has shown that Lang's conjecture (F.4.3.3) would be a consequence of Merel's theorem (F.4.1.1(iii)) and the conjectural lower bound (F.3.4(a)) for the canonical height of nontorsion points on elliptic curves. In particular, (F.4.3.3) is known to be true for elliptic curves with a bounded number of primes dividing $j(E)$ (Silverman [3]), and more generally for elliptic curves with bounded Szpiro ratio (Hindry–Silverman [1]). Similarly, de Diego [1] shows that a universal bound for torsion on abelian varieties and the conjectural height lower bound (F.3.4(b)) would suffice to prove (F.4.3.3(ii)). Over function fields, the elliptic curve bound $\#E(R_{k,S}) \le c^{1+\operatorname{rank} E(k)+\#S}$ is known (Hindry–Silverman [1]); and again over function fields, Buium [3] has proven a bound for the number of rational points $C(k)$ on a curve of genus $g \ge 2$ in terms of g and $\operatorname{rank} \operatorname{Jac}(C)(k)$.

The last quantitative type of estimate we discuss is surprising in its strength.

Question F.4.3.4. *Let k be a number field and let $g \geq 2$ be an integer. Does there exist a bound $B(k,g)$ such that for all curves C/k of genus g, we have*

$$\#C(k) \leq B(k,g)?$$

(One might even ask whether B can be chosen solely in terms of g and the degree $[k:\mathbb{Q}]$.)

Clearly, an affirmative answer to (F.4.3.4) would supersede conjecture (F.4.4.3(ii)), but is there any reason to suspect that such a strong uniform bound should be true? The answer is a conditional yes, based on a fascinating conjecture of Bombieri and Lang. We will discuss the Bombieri–Lang conjecture in more detail below (see Section F.5.2), but briefly it asserts that the k-rational points on a variety of general type are not Zariski dense. Although far from obvious, it can be shown that this statement implies the uniform bound postulated in (F.4.3.4).

Theorem F.4.3.5. (Caporaso–Harris–Mazur [1]) *Assume that the Bombieri–Lang conjecture (F.5.2.1) is true. Then there is a constant $B(k,g)$ such that for all curves C/k of genus g, we have $\#C(k) \leq B(k,g)$.*

We feel obliged to point out that although many mathematicians feel that (F.4.3.5) is good evidence for the existence of the uniform bound for $\#C(k)$, there are others who feel that (F.4.3.5) is primarily evidence for the falsity of the Bombieri–Lang conjecture! We also note that (F.4.3.5) can be extended to points of bounded degree on curves of genus $g \geq 2$ and to integral points on elliptic curves; see Abramovic [1, 2] and Pacelli [1, 2].

F.5. Geometry Governs Arithmetic

Let X/k be a projective variety defined over a number field, say for concreteness defined as a subset of $\mathbb{P}^n$ by a system of homogeneous polynomials

$$F_1(X) = F_2(X) = \cdots = F_r(X) = 0.$$

One of the ultimate goals of Diophantine geometry is to link the geometry of X to its arithmetic. In other words, use algebro-geometric invariants of the complex variety

$$X(\mathbb{C}) = \{x \in \mathbb{P}^n(\mathbb{C}) \mid F_1(x) = \cdots = F_r(x) = 0\}$$

to describe properties of the arithmetic set

$$X(k) = \{x \in \mathbb{P}^n(k) \mid F_1(x) = \cdots = F_r(x) = 0\},$$

or more generally of $X(k')$ for finite extensions of k.

There is a highly developed classification theory for algebraic varieties, and it is tempting to use this geometric theory of classification as the basis of a corresponding Diophantine classification. We will begin with an overview of the geometric classification, especially the notion of Kodaira dimension of a variety. We will then describe several ways in which the geometry of a variety (conjecturally) governs its arithmetic, including a number of beautiful conjectures due to Bombieri, Lang, Vojta, Manin, Batyrev, and others.

Remark. In addition to the algebro-geometric classification of varieties, Lang has suggested that various analytic and differential geometric properties of a variety should influence its arithmetic. We will not have space here to discuss Lang's fascinating conjectures, so we refer the reader to Lang [13] for details.

F.5.1. Kodaira Dimension

Let X be a smooth projective variety. We consider the linear system $|K_X|$ associated to a canonical divisor, and more generally the pluricanonical linear systems $|mK_X|$ attached to multiples of K_X. Assuming that the system is nonempty, we can look at the associated rational map

$$\Phi_{mK_X} : X \dashrightarrow \mathbb{P}^N,$$

which is well-defined up to a linear change of coordinates of $\mathbb{P}^N$.

Definition. Let X be a smooth projective variety as above.

(i) The *plurigenera* of X are the numbers $g_m(X)$ defined by

$$g_m = g_m(X) = \dim L(mK_X) = \ell(mK_X) = h^0(X, \omega_X).$$

(ii) The *Kodaira dimension* of X, denoted by $\kappa(X)$, is the quantity

$$\kappa(X) = \max_{m\geq 1} \dim \Phi_{mK_X}(X);$$

that is, $\kappa(X)$ is the maximal dimension of the image of X under the pluricanonical maps Φ_{mK_X}. If all of the g_m's are zero, the Kodaira dimension is set to $\kappa(X) = -1$ by convention. (Some authors instead prefer to set $\kappa(X) = -\infty$ in this situation.)

The plurigenera $g_m(X)$ and the Kodaira dimension $\kappa(X)$ are birational invariants of X. This follows from Lemma A.1.4.7. It is clear that the Kodaira dimension satisfies $-1 \leq \kappa(X) \leq \dim(X)$.

Examples 5.1.1.

(a) The canonical divisor on projective space is $K_{\mathbb{P}^n} = -(n+1)H$, where H is a hyperplane. Hence $g_m(\mathbb{P}^n) = 0$ for all $m \geq 1$, and $\kappa(\mathbb{P}^n) = -1$.

(b) Let X be a curve of genus g. (Notice that $g = g_1$.)

- If $g = 0$, then $g_m(X) = 0$ for all $m \geq 1$ and $\kappa(X) = -1$.
- If $g = 1$, then $g_m(X) = 1$ for all $m \geq 1$ and $\kappa(X) = 0$.
- If $g \geq 2$, then $g_1(X) = g$, $g_m(X) = (2m-1)(g-1)$ for $m \geq 2$, and hence $\kappa(X) = 1$.

(c) Let X be a smooth complete intersection of dimension $n-r$ in $\mathbb{P}^n$, say the intersection of r hypersurfaces of degrees $d_1, \ldots, d_r$. Then $K_X = (-n-1+d_1+\cdots+d_r)H$, where H is a hyperplane section (see Exercise A.2.7). Hence

$$\kappa(X) = \begin{cases} -1 & \text{if } n+1 > d_1+\cdots+d_r, \\ 0 & \text{if } n+1 = d_1+\cdots+d_r, \\ \dim(X) & \text{if } n+1 < d_1+\cdots+d_r. \end{cases}$$

(d) Let X be a subvariety of an abelian variety A, and let $G_X \subset A$ be its stabilizer. Then (Ueno [1, pages 120–121])

$$\kappa(X) = \dim(X) - \dim G_X.$$

(e) If $K_X = 0$, or more generally if some multiple of K_X is zero, then $\kappa(X) = 0$. This includes abelian varieties, K3 surfaces, Enriques surfaces, and bielliptic surfaces.

The Kodaira dimension of various other varieties may be computed using the next lemma.

Lemma F.5.1.2.

(i) *Let $f : X \dashrightarrow Y$ be a dominant rational map. Then $\kappa(X) \geq \kappa(Y)$.*

(ii) $\kappa(X \times Y) = \begin{cases} -1 & \textit{if } \kappa(X) = -1 \textit{ or } \kappa(Y) = -1, \\ \kappa(X)+\kappa(Y) & \textit{otherwise.} \end{cases}$

(iii) *Let $f : X \to Y$ be an unramified finite covering. Then $\kappa(X) = \kappa(Y)$.*

Proof. (i) The fact that f is dominant means that composition with f induces an injective map $f^* : L(mK_Y) \hookrightarrow L(mK_X)$, so we obtain a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{\Phi_{mK_X}} & \mathbb{P}^{g_m(X)} \\ \downarrow f & & \downarrow \pi \\ Y & \xrightarrow{\Phi_{mK_Y}} & \mathbb{P}^{g_m(Y)} \end{array}$$

where π a linear projection. Thus $\pi \circ \Phi_{mK_X} = \Phi_{mK_Y} \circ f$, from which we deduce that $\dim \Phi_{mK_Y}(Y) \leq \dim \Phi_{mK_X}(X)$.

(ii) Let $p : X \times Y \to X$ and $q : X \times Y \to Y$ be the projections. The canonical divisor of the product is $K_{X\times Y} = p^*K_X + q^*K_Y$, so it is clear

that if $|mK_X| = \emptyset$ or $|mK_Y| = \emptyset$, then also $|mK_{X\times Y}| = \emptyset$. Otherwise, we consider the associated maps, which fit into the commutative diagram

$$\begin{array}{ccc} X \times Y & \xrightarrow{\Phi_{mK_X} \times \Phi_{mK_Y}} & \mathbb{P}^r \times \mathbb{P}^s \\ \big\downarrow{\scriptstyle \Phi_{mK_{X\times Y}}} & & \big\downarrow{\scriptstyle S} \\ \mathbb{P}^t & \xrightarrow{i} & \mathbb{P}^{rs+r+s} \end{array}$$

where S is the Segre embedding and i is a linear embedding. The formula $\kappa(X \times Y) = \kappa(X) + \kappa(Y)$ is clear from this diagram.
(iii) See Iitaka [1, Chapter 10, Theorem 10.9]. □

We now briefly discuss the cases $\kappa = -1$, $\kappa \geq 0$, and $\kappa = \dim(X)$.

Case A. $\kappa(X) = -1$

Note that $\kappa(X) = -1$ is equivalent to $g_m(X) = 0$ for all $m \geq 1$. This will certainly be the case if X is birational to $Y \times \mathbb{P}^1$. More generally, a dominant rational map $\phi : X' \dashrightarrow X$ induces an injection $\phi^* : L(mK_X) \hookrightarrow L(mK_{X'})$, so $g_m(X) \leq g_m(X')$. In particular, if X is dominated by a variety X' satisfying $\kappa(X') = -1$, then necessarily $\kappa(X) = -1$. This motivates the following classical definition.

Definition. A variety X of dimension n is *uniruled* if there exists a variety Y of dimension $n-1$ and a dominant rational map $f : Y \times \mathbb{P}^1 \dashrightarrow X$.

Proposition F.5.1.3. *If X is uniruled, then $\kappa(X) = -1$.*

It is conjectured that the converse is true, that is, $\kappa(X) = -1$ is equivalent to X being uniruled. This is known to be true if $\dim(X) \leq 3$. See Kollar [1] and Debarre [1].

The following provides an important class of varieties with Kodaira dimension -1.

Definition. A *Fano variety* is a smooth projective variety X whose anti-canonical divisor $-K_X$ is ample.

Projective spaces, Grassmannians, and more generally flag varieties are examples of Fano varieties, as are smooth complete intersections of type $(d_1, \ldots, d_r)$ in $\mathbb{P}^n$ with $n+1 > d_1 + \cdots + d_r$. (A *flag variety* is the quotient of a semisimple group by a parabolic subgroup.) It should be noted that the property of being a Fano variety is not birationally invariant. For example, if X is the blowup of $\mathbb{P}^2$ at r points, then X is a Fano variety if and only if $r \leq 8$ and the points are in sufficiently general position. (See Manin [2] for a precise statement and proof.) Fano varieties are covered by rational curves and are uniruled. In fact, one can say much more.

Theorem F.5.1.4. (Mori [1]) *Let X be a Fano variety. For every point $x \in X$ there is a rational curve C such that $x \in C$. Further, the curve C may be chosen to satisfy* $\deg C = -C.K_X \leq \dim X + 1$.

Case B. $\kappa(X) \geq 0$

Consideration of the rational maps $\Phi_{mK_X} : X \dashrightarrow \mathbb{P}^{g_m(x)-1}$ allows one to prove the following result that is fundamental for the classification theory of algebraic varieties.

Theorem F.5.1.5. *Let X be a variety with $\kappa(X) \geq 0$. Then there exists a smooth projective variety Y of dimension $\kappa(X)$, a projective variety X' birational to X, and a surjective morphism $X' \to Y$ whose generic fiber has Kodaira dimension zero.*

In Theorem F.5.1.5, it is not true in general that $\kappa(Y) = \kappa(X)$, see, for example, Exercise F.7.

Example F.5.1.6. If $\kappa(X) = \dim(X) - 1$, then the fibers of the map $X' \to Y$ in (F.5.1.5) are curves of Kodaira dimension 0, hence are elliptic curves. In particular, if X is a surface with $\kappa(X) = 1$, then X is an elliptic surface.

Case C. $\kappa(X) = \dim(X)$

In some sense it is true that the Kodaira dimension of "most" varieties takes on the maximal value $\kappa(X) = \dim(X)$. This prompts the following definition.

Definition. A variety X is a *variety of general type* if $\kappa(X) = \dim(X)$.

Example F.5.1.7. A smooth hypersurface in $\mathbb{P}^n$ of degree $d > n + 1$ is of general type, as is a subvariety of an abelian variety having a finite stabilizer. The term "general type" is classical and not very illuminating. It comes originally from the classification of surfaces:

Kodaira dimension	Types
$\kappa(X) = -1$	rational or ruled
$\kappa(X) = 0$	abelian, bielliptic, K3, or Enriques
$\kappa(X) = 1$	elliptic
$\kappa(X) = 2$	general type ("the others")

Classification of Surfaces

The Kodaira–Parshin surfaces used by Faltings in his proof of the Mordell conjecture are of general type (see Exercise F.8).

F.5.2. The Bombieri–Lang Conjecture

The fundamental Diophantine condition conjecturally satisfied by varieties of general type is the following.

Conjecture F.5.2.1. *Let X be a variety of general type defined over a number field k. Then $X(k)$ is not Zariski dense in X.*

We can make this conjecture even more precise by asserting that except for finitely many points, the Zariski closure of $X(k)$ in X stabilizes as k is enlarged.

Conjecture F.5.2.2. (refined form) *Let X be a variety of general type defined over a number field k. Then there is a dense Zariski open subset U of X such that for all number fields k'/k, the set $U(k')$ is finite.*

Bombieri posed Conjecture F.5.2.1 for surfaces of general type, and Lang (independently) formulated the general conjecture in its refined form (F.5.2.2). In fact, Lang even gave a conjectural description of the exceptional Zariski closed subset. Since we know that projective spaces and abelian varieties may have dense sets of rational points, the exceptional subset must include every image of such varieties. However, we can cover $\mathbb{P}^n$ with projective lines, and the projective line $\mathbb{P}^1$ is the image of an elliptic curve, so the following definition is reasonable.

Definition. Let X be a projective variety. The *special subset* Sp_X of X is the Zariski closure of the union of all images of nontrivial rational maps $A \dashrightarrow X$, where A is an abelian variety.

This allows us to state the final form of the Bombieri–Lang conjecture.

Bombieri–Lang Conjecture F.5.2.3. (final form) *Let X be a variety defined over a number field k, and let $U := X \smallsetminus \mathrm{Sp}_X$. Then $U(k')$ is finite for every finite extension k'/k.*

There are several other ways to formulate Conjecture F.5.2.3 and other possible definitions for the special subset. For example, one may define Sp_X to be the union of all subvarieties of X that are not of general type. It is easy to see that if one takes this definition for Sp_X, then the original Conjecture F.5.2.1 is actually equivalent to the final form (F.5.2.3), provided that the union of subvarieties not of general type inside a variety of general type is a proper closed subset.

There is one further feature noticed by Serge Lang that is worth mentioning. He suggests that being of general type should be analogous to being "pseudo-hyperbolic." More precisely, a variety X is said to be *hyperbolic* if there are no nonconstant holomorphic maps $\mathbb{C} \to X(\mathbb{C})$. Then Lang conjectures that the following three conditions are equivalent:

- $X(k)$ is finite for every number field k.
- Every closed subvariety of X (including X itself) is of general type.
- X is hyperbolic.

He conjectures a similar equivalence on Zariski open subsets, in which case he add the prefix "pseudo" to each condition. See Lang [13] for further details.

Remark 5.2.4. (a) The simplest unknown case of Conjecture F.5.2.1 is a smooth surface of degree 5 in $\mathbb{P}^3$. For example, let $X \subset \mathbb{P}^3$ be the hypersurface given by the equation

$$X_0^5 + X_1^5 + X_2^5 + X_3^5 = 0.$$

It has not been proven that $X(\mathbb{Q})$ consists of a finite number of points beyond those lying on lines and elliptic curves. Note that X does contain several lines, for example $X_0 + X_1 = X_2 + X_3 = 0$, so $X(\mathbb{Q})$ is infinite.
(b) Let $r = \operatorname{rank}(\Omega_X^1)$. If $\Omega_X^r := \bigwedge^r \Omega_X^1$ is ample, then the conjectures predict that $X(k)$ should be finite. Over function fields, this is known to be true by the work of Noguchi [1]. Over number fields it is known only under the additional hypothesis that Ω_X^1 is generated by global sections, in which case the proof reduces to Faltings' theorem on rational points on subvarieties of abelian varieties. (See Moriwaki [1].)
(c) It is natural to ask the complementary question:

> For which varieties X/k does there exist a finite extension k'/k such that $X(k')$ is Zariski dense in X?

This question is still largely unexplored, although it seems reasonable to expect that X has this property if its anticanonical divisor is effective. One might perhaps be bolder and conjecture the same for varieties of Kodaira dimension 0.

However, the answer to the question must depend on more than merely the Kodaira dimension of X, as the following argument shows. Let $X \to C$ be an elliptic surface over a base curve C of genus at least 2. Then the set $X(k')$ is never Zariski dense. On the other hand, if $X \to \mathbb{P}^1$ is an elliptic surface with an infinite group of sections over k', the set $X(k')$ will be Zariski dense. And in both cases, X may have Kodaira dimension 1.

From the previous sections, from (F.1.1.1) in particular, we see that Conjecture F.5.2.3 holds for subvarieties of abelian varieties. Indeed, in that case the special subset Sp_X is the union of translates of nontrivial abelian subvarieties. Although this is essentially the only proved case, there are other sorts of varieties that can reduced to this case. For example, a Kodaira–Parshin surface S is a surface fibered over a curve $S \to C$ such that genus$(C) \geq 2$ and such that the fibers are curves of genus $g' \geq 2$. Then $S(k')$ is finite. Similarly, if the generic fiber is a curve of genus $g' \geq 2$, then $S(k')$ is not Zariski dense. (See Szpiro [2].)

The following proposition allows one to construct additional examples from known examples.

Proposition F.5.2.5. (i) *Let $X \dashrightarrow Y$ be a dominant rational map between varieties of general type. If Conjecture F.5.2.1 is true for Y, then it is true for X (and similarly for Conjecture F.5.2.3).*

(ii) *Let $X \to Y$ be a finite unramified morphism. Then X is of general type if and only if Y is of general type. Hence Conjecture F.5.2.1 (respectively F.5.2.3) is true for X if and only if it is true for Y.*

PROOF. Statement (i) is trivial. The first statement of (ii) follows from Lemma 4.1.1(iii), and the second follows from the Chevalley–Weil theorem (Exercise C.7). □

Remark F.5.2.6. Conjecturally, if X admits a dominant rational map to a variety of general type, then $X(k)$ is never Zariski dense. The following example of Colliot-Thélène, Swinnerton-Dyer, and Skorobogatov [1] (Corollary 3.2) shows that this condition is not necessary.

Let E be a curve of genus 1 with a fixed-point-free involution $\sigma : E \to E$. For example, E could be an elliptic curve and σ translation by a 2-torsion point. Let C be a hyperelliptic curve with hyperelliptic involution $j : C \to C$. Let X be the surface

$$X := (E \times C)/(\sigma \times j).$$

That is, X is the quotient of $E \times C$ by the identification $(x, y) = (\sigma(x), j(y))$. Then one can show that X has no dominant maps to a variety of general type, yet for every number field k, the set $X(k)$ is not Zariski dense in X. Indeed, the latter property is clear, since $(E \times C)(k)$ is not Zariski dense by Faltings' theorem (note that genus$(C) \geq 2$), so the same is true for $X(k)$ because the map $E \times C \to X$ is an unramified finite cover.

One of the most surprising consequences of the Bombieri–Lang conjecture, already mentioned earlier (F.4.3.5) and restated here, is the following uniformity property for rational points on elliptic curves:

> There is a universal upper bound $B = B(g, k)$ such that
> for every curve C/k of genus $g \geq 2$, we have $\#C(k) \leq B$.

The proof, due to Caporaso, Harris, and Mazur [1], is based on the following lemma, which is interesting in its own right.

Lemma F.5.2.7. *Let $f : X \to S$ be a family of curves of genus $g \geq 2$, and let $X_S^n := X \times_S X \times \cdots \times_S X$ be the n-fold fibered product of X over S. Then for sufficiently large n, the variety X_S^n dominates a variety of general type (i.e., there exists a variety W of general type and a dominant rational map $X_S^n \dashrightarrow W$).*

PROOF (that Lemma F.5.2.7 implies uniform boundedness (F.4.3.5)). We apply the lemma to a "universal" family $X \to S$ containing all curves of genus g. From the lemma, X_S^n dominates a variety of general type. Thus,

assuming the Bombieri–Lang conjecture (F.2.1.1), $X_S^n(k)$ is not Zariski dense in X_S^n. Let $U = U_n$ be a nontrivial Zariski open subset of X_S^n such that $U(k) = \emptyset$, define $\pi_j : X_S^{j+1} \to X_S^j$ to be the projection that omits the last factor, and set

$$U_j = \pi_j(U_{j+1}) \qquad \text{and} \qquad Z_j = X_S^j \smallsetminus U_j.$$

For geometric reasons, we have $\#\left(\pi_{j-1}^{-1}(x) \cap Z_j\right) \leq d_{j-1}$ for a constant d_{j-1} independent of x. In particular, for all $x \in U_{n-1}(k)$, the curve $\pi_{n-1}^{-1}(x)$ has at most d_{n-1} points. Going down, we find that for $s \in U_0(k)$, the curve X_s has at most $\max_j\{d_j\}$ points. Since $\dim(S \smallsetminus U_0) < \dim S$, we can repeat the argument on the components of $S \smallsetminus U_0$, so we are done by downward induction on the dimension of S. □

Lemma F.5.2.7 and its application have been generalized by Abramovic and Voloch [2] (see also Abramovic [1]). Abramovic proves in particular that if $f : X \to S$ is a family of smooth varieties of general type, then for n sufficiently large, the fibered product $X_S^n := X \times_S X \times \cdots \times_S X$ dominates a variety of general type.

F.5.3. Vojta's Conjecture

Inspired by Nevanlinna theory and analogies between theorems on value distributions of meromorphic functions and results of Diophantine approximation, Paul Vojta [3] formulated several far-reaching conjectures that remain essentially untouched today. We begin with a brief overview of his insights. Further details may be found in Vojta [3] and Lang [8].

Recall from Section B.8 that there are local height functions $\lambda_{D,v}$ such that if D is a reduced effective divisor D, then intuitively

$$\lambda_{D,v}(P) = -\log(v\text{-adic distance from } P \text{ to } D).$$

Further, the local heights are related to the global height h_D via a sum over places v of k,

$$h_D(P) = \sum_v d_v \lambda_{D,v}(P) \qquad \text{for all } P \notin \operatorname{supp}(D).$$

One of Vojta's insights is that $\lambda_{D,v}$ is the arithmetic analogue of the proximity function in Nevanlinna theory. Indeed, if s is a section of a line bundle on a projective variety X, and if we let $D = \operatorname{div}(s)$ and select a metric $|\cdot|_v$ on the line bundle, then the function $P \mapsto -\log\left|s(P)\right|_v$ is a local height at v.

Let S be a finite set of places of a number field k. The following quantities are counterparts of classical quantities used to study values of meromorphic functions in Nevanlinna theory:

$$m_S(D,P) = \sum_{v\in S} d_v \lambda_{D,v}(P) = \sum_{v\in S} -d_v \log|s(P)|_v,$$

$$N_S(D,P) = \sum_{v\notin S} d_v \lambda_{D,v}(P) = \sum_{v\notin S} -d_v \log|s(P)|_v.$$

Notice that by definition

$$h_D(P) = N_S(D,P) + m_S(D,P).$$

The global height function $h_D(\,\cdot\,)$ is the arithmetic analogue of the characteristic function from classical Nevanlinna theory, and similarly $m_S(D,\,\cdot\,)$ is the arithmetic analogue of the proximity function and $N_S(D,\,\cdot\,)$ is the arithmetic analogue of the counting function.

Definition. If X/k is a projective variety, then the set of rational points $X(k)$ of X is a precisely defined set; but if R is a subring of k and U is an affine open subset of X, then the set of integral points $U(R)$ of U is ambiguous, since it depends on choosing a particular collection of affine coordinate functions $x_1,\dots,x_n$ on U. Thus integrality is really a property of a *set* of points $\mathcal{P} \subset X(k)$, rather than a property of individual points. We say that $\mathcal{P}$ is a *set of integral points of* U if there exist affine coordinates $x_1,\dots,x_n$ on U such that $x_i(P) \in R$ for all $1 \le i \le n$ and all $P \in \mathcal{P}$. Note that according to this definition, any finite set is automatically integral, so generally one studies whether or not there exist infinite integral sets.

We can use Vojta's ideas to formulate a more general notion of integrality. Thus let k be a number field, let $R = R_S$ be the ring of S-integers of k, and let $D = X \smallsetminus U$ be the divisor at infinity. A subset $\mathcal{P}$ of $U(k)$ will be called *S-integral* if there is a constant c such that

$$m_S(D,P) \ge h_D(P) - c \qquad \text{for all } P \in \mathcal{P}.$$

Intuitively, the inequality $m_S(D,P) \ge h_D(P) - c$ means that virtually all of the v-adic closeness of P to D occurs for the places v in S. More generally, we say that a subset $\mathcal{P}$ of $U(k)$ is *quasi-S-integral* if there is are constants c and $\varepsilon > 0$ such that

$$m_S(D,P) \ge \varepsilon h_D(P) - c \qquad \text{for all } P \in \mathcal{P}.$$

Example F.5.3.1. Let $X = \mathbb{P}^1$ and $D_\alpha = (\alpha)$, where $\alpha \in \bar{k}$ is an algebraic number. The (arithmetic) defect of α is the quantity

$$\delta(\alpha) = \liminf_{a\in\mathbb{P}^1(k),\ h(a)\to\infty} \frac{m_S(D_\alpha, a)}{h_{D_\alpha}(a)}.$$

Then one can show (Vojta [3]) that Roth's theorem is equivalent to the inequality $\delta(\alpha) \le 2$, which is a perfect analogue of the defect relation of Nevanlinna theory.

Based on this example and further analogies with higher-dimensional results in Nevanlinna theory, Vojta formulated the following conjecture:

Conjecture F.5.3.2. (Vojta) *Let S be a finite set of places of a number field k. Let X/k be a smooth projective variety, let E be an ample divisor on X, and let D be an effective reduced divisor on X having only normal crossings. Then for every $\varepsilon > 0$ there exists a proper closed subset $Z = Z(\varepsilon, E, D, k, S)$ such that*

$$m_S(D,P) + h_{K_X}(P) \le \varepsilon h_E(P) + O_\varepsilon(1) \qquad \text{for all } P \in (X \smallsetminus Z)(k).$$

Remark F.5.3.3.
(a) The set Z in Vojta's conjecture is called the *exceptional subset*. A strengthened version of Vojta's conjecture says that there is a geometric exceptional set $Z_{\text{geom}} = Z_{\text{geom}}(\varepsilon, E, D)$ such that for all extension fields k'/k and finite sets of places S' of k',

$$\{P \in (X \smallsetminus Z_{\text{geom}})(k') \mid m_{S'}(D,P) + h_{K_X}(P) \ge \varepsilon h_E(P) + O_\varepsilon(1)\}$$

is a finite set. In other words, aside from a finite number of exceptional points, the exceptional set may be chosen independently of the field k and the set of places S.
(b) If X is a variety of general type, then Vojta's conjecture with $D = 0$, $E = K_X$, and $\varepsilon = \frac{1}{2}$ says that $h_{K_X}(P) \le O(1)$ for $P \in (X \smallsetminus Z)(k)$. Since K_X is almost ample, this implies that $X(k)$ is not Zariski dense in X. Thus the Bombieri–Lang conjecture (F.5.2.1) is a special case of Vojta's conjecture.
(c) Let $X \subset \mathbb{P}^n$ be a smooth projective variety with $K_X = 0$, for example a K3 surface or an abelian variety, let $U = X \cap \mathbb{A}^n$ be an affine subvariety of X, and let $D = X \smallsetminus U$ be the reduced divisor consisting of the "points at infinity" on X. Notice that D is necessarily ample. Vojta's conjecture (with $E = D$) says that $m_S(D,P) \le \varepsilon h_D(P) + O_\varepsilon(1)$. It follows that a quasi-S-integral subset of $U(k)$ is contained in a proper Zariski closed subset of X. In the case of abelian varieties, Faltings [2] has proven that for an ample divisor D with normal crossings, we have $m_S(D,x) \le \varepsilon h(x)$ outside a closed subset (hence everywhere by induction).
(d) Vojta's conjecture is known for curves. Thus for curves of genus 0 it is equivalent to Roth's theorem, and for curves of genus $g \ge 2$ it is a consequence of Faltings' theorem. As for curves of genus 1, it is equivalent to the results (D.8.3) and (D.8.4) used to prove Siegel's theorem.
(e) We note that the requirement in Vojta's conjecture that D have only normal crossings is necessary, since it is easy to produce counterexamples if this condition is dropped.

One other case in which Vojta's conjecture is known is the deep generalization of Roth's theorem proven by Wolfgang Schmidt in 1970.

Subspace Theorem F.5.3.4. (Schmidt [2, 4]) *Let*

$$L_1, \dots, L_m \in \bar{\mathbb{Q}}[x_1, \dots, x_n]$$

be linear forms in general position, i.e., such that any subset of at most $\min\{m,n\}$ *of the forms are linearly independent. Then there exist finitely many proper linear subspaces* $T_1, \dots, T_r \subset \mathbb{C}^n$ *such that*

$$\{x \in \mathbb{Z}^n \mid |x|^\varepsilon \cdot |L_1(x) \cdots L_m(x)| \le 1\} \subset \bigcup_{i=1}^r T_i.$$

To see that Roth's theorem follows from (F.5.3.4), take $m = n = 2$, $L_1(x,y) = x - \alpha y$, and $L_2(x,y) = y$.

Schmidt's subspace theorem has been generalized by Schlickewei [1] to number fields and to include more than one absolute value. The appropriate inequality for the general statement has the form

$$\sum_{v \in S} \log \max_j \prod_i \left| \frac{x_j}{L_i(x_0, \dots, x_n)} \right|_v \le (n+1+e)h(x) \qquad \text{for } x \in (\mathbb{P}^n \setminus T)(k),$$

where T is a union of linear subspaces of $\mathbb{P}^n$. If we now observe that $K_{\mathbb{P}^n}$ is $-(n+1)$ times a hyperplane, so $h_{K_{\mathbb{P}^n}} = -(n+1)h + O(1)$, then this inequality may be written in the form

$$\sum_{v \in S} \lambda_{D,v}(x) + h_{K_{\mathbb{P}^n}}(x) \le \varepsilon h(x) + O(1),$$

which is precisely Vojta's inequality. Thus Schmidt's theorem is a special (highly nontrivial) case of Vojta's conjecture, with the addition that the exceptional set Z is specified as a collection of linear subspaces. Note that the "general position" requirement in Schmidt's subspace theorem is exactly the "normal crossings" condition in Vojta's conjecture.

Remark F.5.3.5. Let $Q(x_0, \dots, x_n)$ be a homogeneous form of degree d with $d \ge n+2$, and let D be the divisor in $X = \mathbb{P}^n$ defined by $Q = 0$. Assuming that D has only normal crossings, we see that Vojta's conjecture predicts that for any integer $b \ne 0$, the set

$$\{x \in \mathbb{Z}^{n+1} \mid Q(x) = b\}$$

lies in a proper Zariski closed subset of $\mathbb{P}^n(\mathbb{Q})$. Note that if $Q = L_1 L_2 \cdots L_d$ is the product of d linear forms in general position, then this follows immediately from Schmidt's subspace theorem (F.5.3.4); but in general it is still an open question. Indeed, it does not appear to be known whether the integer solutions to the specific equation

$$x^5 + y^5 + z^5 + w^5 = 1$$

are Zariski dense in $\mathbb{P}^3(\mathbb{Q})$.

This example also clearly shows the necessity of the normal crossing condition. Thus if we take

$$Q(x_0, \ldots, x_n) = x_0^{d-1} x_n + Q'(x_0, \ldots, x_{n-1}),$$

then it is easy to see that the S-integral solutions to $Q(x) = 1$ will be Zariski dense in $\mathbb{P}^n(k)$, provided that the unit group R_S^* is infinite. (Choose $x_0 \in R_S^*$ and $x_1, \ldots, x_{n-1} \in R_S$ arbitrarily and solve for x_n.) See Exercise 9 or Vojta's original account (Vojta [3]) for further examples.

A convenient reformulation of certain cases of Vojta's conjecture uses the following generalization of the notion of a variety of general type.

Definition. A quasi-projective variety U is of *log general type* if it can be written $U = X \smallsetminus D$ with X projective, D an effective divisor with normal crossings, and $K_X + D$ almost ample on X.

Conjecture F.5.3.6. (Lang–Vojta) *Let U/k be a variety of log general type. Then any set of S-integral points on U is contained in a proper Zariski closed subset.*

Conjecture F.5.3.6 is true for curves by Siegel's theorem, since $K_X + D$ is ample in exactly the following three situations: (i) $g \geq 2$; (ii) $g = 1$ and $\deg D \geq 1$; (iii) $g = 0$ and $\deg D \geq 3$. More generally, (F.5.3.6) says that there are finitely many integral points in an affine open subset of an abelian variety, a result (as already mentioned) proven by Faltings [2].

Example F.5.3.7. Let $\mathcal{A}_{g,N}$ be the moduli space of principally polarized abelian varieties of dimension g with level-N structure. For sufficiently large values of N, the points in $\mathcal{A}_{g,N}(R_{k,S})$ essentially correspond to isomorphism classes of principally polarized abelian varieties of dimension g with level-N structure defined over k and having good reduction outside of S. Faltings [1] has proven that this set is finite, a result originally conjectured by Shafarevich. This amounts to showing that the set of S-integral points with respect to the divisor D is finite, where D is the divisor at infinity $D = \overline{\mathcal{A}}_{g,N} \smallsetminus \mathcal{A}_{g,N}$ on a "nice" compactification of the moduli space. Faltings' result is compatible with Conjecture F.5.3.6 in the sense that it is known that $\mathcal{A}_{g,N}$ is of log general type for sufficiently large N. A similar remark applies to the moduli space $\mathcal{M}_{g,N}$ of curves of genus g with level-N structure.

Vojta also proposes a conjecture involving all algebraic points on a variety, not only the points rational over a particular field. In order to state the general conjecture, we introduce the absolute logarithmic discriminant

$$d_k(P) = \frac{1}{[k(P) : \mathbb{Q}]} \log\left|\operatorname{Disc}(k(P)/\mathbb{Q})\right|,$$

and we generalize the definitions of $m_S(D,P)$ and $N_S(D,P)$ to $P \in X(\bar{\mathbb{Q}})$ using the extension of local heights to algebraic points described in (B.8.3),

$$\lambda_{D,v}(P) = \frac{1}{[k(P):k]} \sum_{w \in M_{k(P)},\ w|v} [k(P)_w : k_v]\lambda_{D,w}(P).$$

Conjecture F.5.3.8. (Vojta) *Let S be a finite set of places of a number field k. Let X/k be a smooth projective variety, let E be an ample divisor on X, and let D be an effective reduced divisor on X having only normal crossings. Then for every $\varepsilon > 0$ and $r \geq 1$ there exists a proper closed subset $Z = Z(\varepsilon, r, E, D, k, S)$ such that*

$$m_S(D,P) + h_{K_X}(P) \leq d_k(P) + \varepsilon h_E(P) + O_\varepsilon(1)$$

for all $P \in (X \smallsetminus Z)(\bar{k})$ satisfying $[k(P):k] \leq r$.

A fascinating aspect of Vojta's conjectures (F.5.3.2) and (F.5.3.8) is that they seem to contain virtually all Diophantine statements that are currently proven or conjectured, by which we mean statements asserting that certain arithmetically defined sets are "small". For example, Vojta's conjectures imply the *abc* conjecture (see Exercise F.11). It should also be emphasized that it was Vojta's philosophy of seeking analogies between Nevanlinna theory and Diophantine approximation that led to his proof of Mordell's conjecture (about eight years after Faltings' initial proof).

F.5.4. Varieties Whose Rational Points Are Dense

The previous sections have dealt principally with the question of when sets of rational or integer points fail to be Zariski dense. In this final section we will discuss the case that $X(k)$ is Zariski dense in X and consider ways to measure that density. For example, if E/k is an elliptic curve of infinite rank, then $E(k)$ is Zariski dense in E, but one feels that somehow it is "less dense" than, say, $\mathbb{P}^1(k)$ is in $\mathbb{P}^1$.

One way to measure the density is to fix a height H on X associated to an ample divisor D and consider the behavior of the counting function

$$N(X/k, D, B) = \#\{x \in X(k) \mid H(x) \leq B\} \quad \text{as } B \to \infty.$$

Notational Convention. It turns out that the asymptotic formulas for counting functions take a simpler form if one uses the height with respect to a particular field, rather than using the normalized height. So for the remainder of this section we will fix a number field k and use the notation $H = H_k$.

Example F.5.4.1. Schanuel's theorem (B.6.2) (Schanuel [1]) says that there is a constant $c = c(n,k)$ such that $N(\mathbb{P}^n/k, D, B) \sim cB^{n+1}$, where D is a hyperplane. Similarly, if A/k is an abelian variety with $r = \operatorname{rank} A(k)$, then there is a constant $c = c(A,k,D)$ such that

$$N(A/k, D, B) \sim c(\log B)^{r/2}$$

for any ample divisor D. This is Theorem B.6.3.

We begin with a geometric discussion about divisor classes. Recall that a *closed cone* in a real vector space V is a subset $\mathcal{C} \subset V$ with the property that $x \in \mathcal{C}$ implies that $tx \in \mathcal{C}$ for all real numbers $t \geq 0$. A cone is said to be *convex* if $x, y \in \mathcal{C}$ implies that $x + y \in \mathcal{C}$.

Definition. Let X be a variety and $\mathrm{NS}(X)$ its Néron–Severi group. The *effective cone of* X, denoted by $\mathrm{NS}_{\mathrm{eff}}(X)$, is the closed cone in $\mathrm{NS}(X) \otimes \mathbb{R}$ generated by the classes of effective divisors. The *ample cone of* X, denoted by $\mathrm{NS}_+(X)$, is the closed cone in $\mathrm{NS}(X) \otimes \mathbb{R}$ generated by the classes of ample divisors.

Clearly, $\mathrm{NS}_+(X) \subset \mathrm{NS}_{\mathrm{eff}}(X)$. We also observe that the property $\mathrm{Pic}^0(X) = 0$ is equivalent to $\mathrm{Alb}(X) = 0$ (i.e., X admits no nonconstant maps to abelian varieties), in which case $\mathrm{Pic}(X) = \mathrm{NS}(X)$. This occurs, for example, for all Fano varieties, and more generally for any variety that is covered by rational curves.

Definition. Let D be an ample divisor on a normal projective variety X. The *Nevanlinna invariant of* D is the number

$$\alpha(D) := \inf\{r \in \mathbb{Q} \mid K_X + rD \in \mathrm{NS}_{\mathrm{eff}}(X)\}.$$

For simplicity we will restrict attention to smooth varieties, but most of what we say will apply to normal varieties. Further, we note that the counting function on a singular variety is more or less equal to the counting function on its normalization; see Exercise F.18.

Remark F.5.4.2.

(i) The invariant $\alpha(D)$ is usually defined in Nevanlinna theory as the number

$$\inf\left\{\frac{p}{q} \;\middle|\; q > 0 \text{ and } pD + qK_X \text{ is almost ample}\right\}.$$

The two definitions coincide if D is ample.

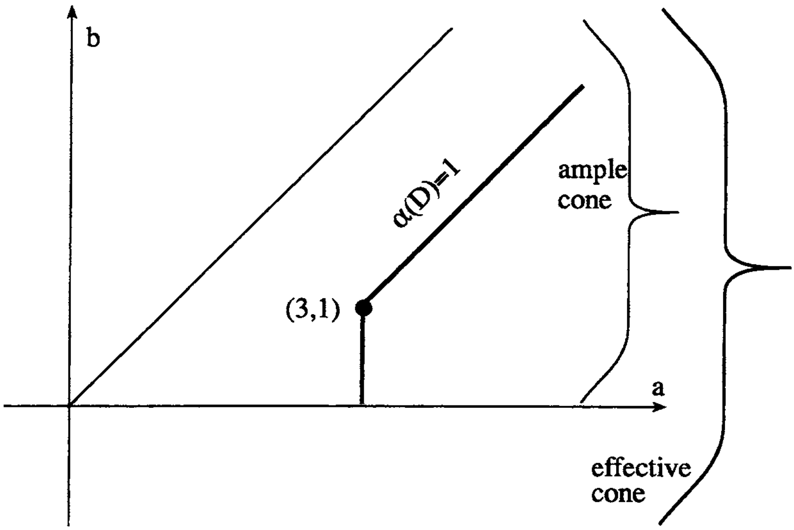
If K_X is not in $\mathrm{NS}_{\mathrm{eff}}(X)$, then $\alpha(D)$ is the real number such that $\alpha(D)[D] \in -[K_X] + \partial\,\mathrm{NS}_{\mathrm{eff}}(X)$, where $\partial\,\mathrm{NS}_{\mathrm{eff}}(X)$ denotes the boundary of the effective cone $\mathrm{NS}_{\mathrm{eff}}(X)$. It is expected that $\alpha(D)$ is always a rational number. We also observe that the Nevanlinna invariant is implicitly present in Vojta's conjecture.

(ii) The Nevanlinna invariant has the following properties:
- Let $f : X \to Y$ be a finite map. Then $\alpha_X(f^*D) \le \alpha_Y(D)$, with equality if f is unramified.
- If $D \ge D'$, then $\alpha(D) \le \alpha(D')$.
- The Nevanlinna invariant is inverse linear: $\alpha(mD) = \frac{1}{m}\alpha(D)$.
- Additivity of the Nevanlinna invariant is more complicated. For example, it is true that
$$\frac{1}{\alpha(D)} + \frac{1}{\alpha(D')} \le \frac{1}{\alpha(D+D')},$$
and hence $\alpha(D+D') \le \max\{\alpha(D), \alpha(D')\}/2$ (Exercise F.13).

Example F.5.4.3. We illustrate the preceding ideas by explicitly computing the relevant quantities for the variety X that is the blowup $X \to \mathbb{P}^2$ of the projective plane at a point. Let L be the pullback to X of a generic line in $\mathbb{P}^2$, and let E be the exceptional divisor (i.e., the line on X that replaces the blown-up point). Then one knows that $\operatorname{Pic}(X) = \operatorname{NS}(X)$ is a free group of rank 2 generated by L and E. Further, the canonical divisor of X is given by $K_X = -3L + E$. A divisor $D = aL - bE$ on X is ample if and only if $a > b > 0$, while the (closed) effective cone is determined by the condition $a \ge \max\{0, b\}$. A short computation shows that the Nevanlinna invariant of $aL - bE$ is
$$\alpha(aL - bE) = \max\left\{\frac{3}{a}, \frac{2}{a-b}\right\},$$
and that the divisor $aL - bE$ is proportional to K_X if and only if $a = 3b$. These results are illustrated in Figure F.1 below. Now let $U \subset X$ be the complement of E. Then it is also not difficult to show (see Exercise F.14 for a more general result) that
$$N(U/k, D, B) = \begin{cases} cB^{\alpha(D)} & \text{if } D \text{ is not proportional to } K_X, \\ cB^{\alpha(D)} \log B & \text{if } D \text{ is proportional to } K_X. \end{cases}$$

Let X/k be a smooth projective variety as usual, and let $U \subset X$ be an open subset (possibly $U = X$ itself). Rather than studying the asymptotic behavior of the counting function $N(U/k, D, B)$, we can use ideas from analytic number theory by introducing the *height zeta function*
$$Z(U/k, D; s) = Z_D(s) = \sum_{x \in U(k)} H_D(x)^{-s}.$$
Here we assume that D is ample and that the Weil height H_D is chosen to satisfy $H_D(x) \ge 1$ for all x. (Alternatively, we could discard a finite number of points from the sum.) The goal is to describe the analytic behavior of $Z_D(s)$ in terms of the geometric properties of X, U, and D.

The zeta function $Z_D(s)$ is a (generalized) Dirichlet series, so it is convergent on some half-plane $\operatorname{Re}(s) > b$. By convention, we set $b = -\infty$ if $Z_D(s)$ is convergent on all of $\mathbb{C}$, which will occur if $U(k)$ is finite. The *abscissa of convergence of* $Z_D(s)$ is the infimum over all b such that $Z_D(s)$ converges on $\operatorname{Re}(s) > b$.



The ample cone, effective cone, and polyhedron $\alpha(D) = 1$ for $D = aL - bE$
Figure F.1

Remark F.5.4.4. Intuitively, the zeta function $Z_D(s)$ should have a pole of order t at its abscissa of convergence β if and only if $N(U/k, D, B) \sim cB^{\beta}(\log B)^{t-1}$. The following lemma makes this idea more precise.

Lemma F.5.4.5. *Let S be a set, and let $H : S \to [1, \infty)$ be a function with the property that S has only finitely many elements with bounded H. Define a counting function and a zeta function in the natural way,*

$$N(S, H, B) = \#\{x \in S \mid H(x) \le B\} \quad \textit{and} \quad Z(S, H; s) = \sum_{x \in S} H(x)^{-s}.$$

Let a be the abscissa of convergence of $Z(S, H; s)$.

(i) *If there is a $c \ge 0$ such that*

$$N(S, H, B) \sim cB^{a}(\log B)^{t-1} \quad \textit{as } B \to \infty, \tag{$*$}$$

then

$$\lim_{s \to a+} (s - a)^t Z(S, H; s) = c\Gamma(t)a. \tag{$**$}$$

(ii) *Assume that the function $Z(S, H; s)$ extends to a holomorphic function at all points on the line* $\mathrm{Re}(s) = a$*, except for a pole at $s = a$. Then* $(**)$ *implies* $(*)$.

Proof. See, for example, Tenenbaum [1, Theorems 2 and 15, Section II.7]. □

We observe that the abscissa of convergence of $Z(S,H;s)$ is equal to $\limsup B^{-1}\log N(S,H,B)$, provided that S is infinite; it is clearly equal to $-\infty$ if S is finite.

There are few general results on the abscissa of convergence for the height zeta function $Z(U/k,D;s)$ of a variety. We will just quote the following result.

Theorem F.5.4.6. Pila [1] *Let X be a variety of dimension n and degree d in $\mathbb{P}^N$. Then for any $\varepsilon > 0$ there is a constant $c = c(n,d,\varepsilon)$ such that*

$$N(X/\mathbb{Q},B) \le cB^{n+\frac{1}{d}+\epsilon}.$$

(Note that c depends only on the dimension and degree of X.)

Although quite weak, Theorem F.5.4.6 is nontrivial if $d > 1$, since the trivial estimate would be $N(\mathbb{P}^n/\mathbb{Q},B) \sim c'B^{n+1}$.

We are now ready to formulate a fundamental conjecture for varieties with a dense set of rational points. This conjecture relates the geometrically defined Nevanlinna invariant to an arithmetically defined abscissa of convergence.

Conjecture F.5.4.7. (Batyrev–Manin [1]) *Let X be smooth projective variety, and let D be an ample divisor on X. For any Zariski open subset of U, let $\beta(U/k,D)$ be the abscissa of convergence of the height zeta function $Z(U/k,D;s)$.*

(i) *For every $\varepsilon > 0$ there exists a dense open subset U of X such that*

$$\beta(U/k,D) \le \alpha(D) + \varepsilon.$$

(ii) *Assume that $K_X \notin NS_{\text{eff}}$, and hence that $\alpha(D) > 0$ for every ample divisor D. Then for all sufficiently large number fields k' and all sufficiently small dense open subsets $U \subset X$, we have*

$$\beta(U/k',D) = \alpha(D).$$

The Nevanlinna invariant $\alpha(D)$ and the height abscissa $\beta(U/k,D)$ obey many of the same formal rules; see Exercise F.13.

Conjecture F.5.4.7 gives the most precise information for Fano varieties, or more generally for varieties whose canonical divisor is not effective. We will discuss these varieties further below. First we indicate briefly what the conjecture says for other sorts of varieties.

For example, if K_X is almost ample, then there is an effective divisor E such that $D = K_X - E$ is ample, so $\alpha(D) = -1$. It follows from (F.5.4.7(i)) that $\beta(U/k,D) < 0$ on some open subset U of X, which means that $U(k)$ is finite. Hence the Batyrev–Manin conjecture (F.5.4.7) is equivalent to the

Bombieri–Lang conjecture that rational points on varieties of general type are not dense.

Another interesting class of examples are those for which the canonical class K_X is trivial in $\mathrm{NS}(X) \otimes \mathbb{R}$. Then $\alpha(D) = 0$ for all ample D, so (F.5.4.7(i)) says that for any $\varepsilon > 0$ there is an open subset U_ε such that $N(U_\varepsilon/k, D, B) \ll B^\varepsilon$. If $X = A$ is an abelian variety, we know that this is true in the more precise form

$$N(A/k, D, B) = c(\log B)^{r/2} + O\left((\log B)^{(r-1)/2}\right)$$

with $r = \operatorname{rank} A(k)$; see Theorem B.6.3.

Now consider the case of a K3 surface or an Enriques surface X. Then the only curves on X with at least B^ε points of height $H_D(x) \leq B$ are rational curves C satisfying $C \cdot D \leq 2/\varepsilon$. Hence in this situation the conjecture reduces to the following.

Conjecture F.5.4.8. (Batyrev–Manin) *Let X/k be a K3 surface or an Enriques surface. For any $\varepsilon > 0$ and any ample divisor D on X, let Z_ε be the (finite) union of rational curves $C \subset X$ defined over k and satisfying $C.D \leq 2/\varepsilon$. Let $U_\varepsilon = X \smallsetminus Z_\varepsilon$ be the complement of Z_ε. Then*

$$N(U_\varepsilon/k, D, B) \ll B^\varepsilon \quad \text{as } B \to \infty.$$

Remarks F.5.4.9.

(a) There are no surfaces for which (F.5.4.8) is known to be true for all number fields, but see Billard [1] for some partial results on K3 surfaces of type $(2,2,2)$ in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$. See also Silverman [6] for height estimates on certain K3 surfaces in $\mathbb{P}^2 \times \mathbb{P}^2$.

(b) Batyrev and Manin ventured a refinement of their conjecture, which in the case of a Fano variety X (i.e., $-K_X$ ample) says that there is an open subset $U \subset X$ such that

$$N(U/k, -K_X, B) \sim cB(\log B)^{\operatorname{rank} \operatorname{Pic}(X/k)-1}.$$

Although this formula is correct in many cases, Batyrev and Tschinkel [1] have given examples in which the exponent on the $\log B$ is incorrect. See (F.5.4.10(i)) for details.

(c) The Batyrev–Manin conjecture (F.5.4.7) gives a geometric interpretation for the first pole of $Z(U/k, D; s)$. We will briefly discuss below the computation, in geometric terms, of the order of that pole.

(d) The necessity of taking an open subvariety in the Batyrev–Manin conjecture (F.5.4.7) and of augmenting the field in (F.5.4.7(ii)) is obvious. For a classical example, let X be a smooth cubic surface in $\mathbb{P}^3$, so X is embedded by $-K_X$. Then $\alpha(-K_X) = 1$, but X contains 27 lines, and if any of

those lines is rational over k, then they will contain cB^2 points of height less than B. Thus it is certainly necessary to discard all of the k-rational lines. Similarly, the example of a conic curve with no k-rational points shows that it may be necessary to extend the ground field.

(e) Let $X \to Y$ be a finite unramified cover, and let D be an ample divisor with $\alpha(D) = 0$. Then it is not hard to show using the Chevalley–Weil theorem (Exercise C.7) that the Batyrev–Manin conjecture (F.5.4.7) (for all number fields) on X is equivalent to the conjecture on Y. See Morita–Sato [1]. For example, this shows that (F.5.4.7) is true for bielliptic surfaces, since they admit an unramified cover by an abelian surface. Similarly, the conjecture for elliptic K3 surfaces is equivalent to the conjecture for Enriques surfaces, since the former are unramified double covers of the latter.

(f) For fibrations, there are a few partial results. Call [1] and Billard [2] give estimates for elliptic surfaces, and Billard [3] proves the Batyrev–Manin conjecture (F.5.4.7) for many (but not all) ample divisors on a rational ruled surface.

A natural class of varieties on which to test the conjectures of this section are Fano varieties. We refer the reader to Manin [2] and Manin–Tsfasman [1] for a geometric discussion of Fano varieties. Of particular interest are Fano surfaces, also called *Del Pezzo surfaces*. These surfaces fit into 10 families, namely $\mathbb{P}^2$ blown up at r points in general position with $0 \le r \le 8$, plus the product $\mathbb{P}^1 \times \mathbb{P}^1$. Similarly, Fano 3-folds fit into 104 families, and in general Fano varieties of dimension n lie in a finite number of families (see Debarre [1]). We now describe some examples for which it is known that

$$N(X/k, D, B) \gg\ll B^{\alpha}(\log B)^{t-1}.$$

Fano Examples F.5.4.10.

(a) We start with an easy example that generalizes Schanuel's theorem. Let

$$X = \prod_{i=1}^{m} \mathbb{P}^{n_i} \qquad \text{and} \qquad D = \sum_{i=1}^{m} p_i^*(d_i H_i),$$

where H_i is a hyperplane in $\mathbb{P}^{n_i}$ and p_i is the projection onto the ith factor of X. Let t be the number of indices for which the quantity $(n_i + 1)/d_i$ is maximized, and let α be that maximal value of $(n_i + 1)/d_i$. Then one can show (Exercise F.15) that

$$N(X, H_D, B) \sim cB^{\alpha}(\log B)^{t-1}.$$

(b) An old result of Birch [1] deals with a smooth complete intersection $X/\mathbb{Q}$ in $\mathbb{P}^n$ defined by r homogeneous polynomials of degree d. If n

is sufficiently large, specifically $n > r(r+1)(d-1)2^{d-1}$, then Birch proves that $N(X/\mathbb{Q}, L, B) \sim cB^{n+1-rd}$, where L is a hyperplane section. Since $K_X = -(n+1-rd)L$, we see that X is a Fano variety and that Birch's result becomes $N(X/\mathbb{Q}, -K_X, B) \sim cB$. Birch also shows that $c \neq 0$ if and only if $X(\mathbb{Q}_v) \neq \emptyset$ for all places v, thereby showing that X satisfies the Hasse principle. The proof is via the circle method. Heath-Brown [1] and Hooley [1, 2, 3] have refined the method to prove the same estimate for smooth cubic hypersurfaces in $\mathbb{P}^n$ for $n \geq 8$. In the same vein, we mention the estimate $N(U/\mathbb{Q}, L, B) \gg\ll B^2(\log B)^5$ proven by Vaughan and Wooley [1] for an open subset of a certain singular cubic threefold in $\mathbb{P}^6$, and work of Heath-Brown [2] on quadrics.

(c) Batyrev and Manin [1] show that if X is Fano and if (F.5.4.7(ii)) is true for $D = -K_X$ (i.e., if $\alpha(-K_X) = \beta_{U,k'}(-K_X)$), then (F.5.4.7(i)) is true for for all ample divisors on X. In the case that X is $\mathbb{P}^2$ blown up at $r \leq 3$ rational points, so $\operatorname{rank}\operatorname{Pic}(X) = 1 + r$, they prove the full conjecture $N(U, -K_X, B) \sim cB(\log B)^r$, where U is X with the exceptional lines removed. For $\mathbb{P}^2$ blown up at 4 points, Manin and Tschinkel [1] prove the weaker estimate

$$B(\log B)^3 \ll N(U, -K_X, B) \ll B(\log B)^6.$$

In particular, this implies that $\beta_U(D) = \alpha(D)$ for split Del Pezzo surfaces $\mathbb{P}^2$ blown up at $r \leq 4$ points.

(d) Manin [3] gives explicit estimates for $N(U/k, D, B)$ for an open subset of $\mathbb{P}^n$ blown up along a linear subspace $\mathbb{P}^m$ (with $m \leq n-2$), generalizing the estimates (F.5.4.3) for $\mathbb{P}^2$ blown up at a point. See Exercise F.14.

(e) Franke, Manin, and Tschinkel [1] prove one of the most general known results, namely that the Batyrev–Manin conjecture (F.5.4.7), in its refined form (F.5.4.9(b)), is true for homogeneous spaces. Precisely, let G be a semisimple algebraic group, let P be a parabolic subgroup of G, and let X be the quotient variety $X = P\backslash G$. Then X is a projective variety (in fact, it is a Fano variety). Assuming that $X(k) \neq \emptyset$, they prove the

$$N(X/k, -K_X, B) \sim cB(\log B)^{\operatorname{rank}\operatorname{Pic}(X)-1}.$$

Notice that there is no exceptional set. The essential ingredient in the analytic proof is showing that for a suitable normalization of the height, the zeta function $Z(X/k, -K_X; s)$ is closely related to an Eisenstein-Langlands L-series.

(f) The result of Franke, Manin, and Tschinkel [1] covers in particular the case of Grassmannian varieties, and more generally flag varieties. These were treated independently by Thunder [1, 2] using more elementary counting techniques. Thunder gives explicit formulas for the constant c and explicit error estimates, both of which could in principle be derived via the Eisenstein–Langlands L-series approach.

(g) Batyrev and Tschinkel [2] prove the refined conjecture (F.5.4.9(b)) for *toric varieties.* These are smooth projective varieties X that contain an open subset U isomorphic to a torus (i.e., $U \cong \mathbb{G}_m^s$ over $\bar{k}$) and with the property that the group law $U \times U \to U$ extends to a morphism $U \times X \to X$. In other words, there is an algebraic group action of U on X. Note that multiples of the anticanonical bundle $-K_X$ give embeddings only of U, not necessarily embeddings of all of X. For example, $(\mathbb{P}^1)^s$ is a toric variety, as is $\mathbb{P}^2$ blown up at 3 points.
(h) In a paper with a strongly adelic flavor, Peyre [1] pushes these ideas further. He also defines an explicit constant $\mathcal{C}(X, -K_X)$ (depending on the choice of height function H_{-K_X}) and conjectures that

$$N(U, -K_X, B) \sim \mathcal{C}(X, -K_X) B (\log B)^{t-1}.$$

Peyre verifies the value of $\mathcal{C}(X, -K_X)$ for the examples in his paper, for Birch's examples (b), and for Thunder's examples (f). However, it appears that for toric varieties Peyre's constant $\mathcal{C}(X, -K_X)$ requires an extra factor that can be interpreted as the order of a Brauer group (see Batyrev–Tschinkel [2]).
(i) As mentioned earlier, Batyrev and Tschinkel [1] have shown that the refined conjecture (F.5.4.9(b)) is not true in general. Their counterexample is a threefold X of bidegree $(1,3)$ in $\mathbb{P}^1 \times \mathbb{P}^3$. Ignoring some technical difficulties, the underlying idea is quite simple. The anticanonical divisor is $-K_X = \mathcal{O}(1,1)$, and $\operatorname{Pic}(X)$ has rank 2, since it is isomorphic to $\operatorname{Pic}(\mathbb{P}^1 \times \mathbb{P}^3)$. Hence the refined Batyrev–Manin conjecture predicts $N(U, -K_X, B) \sim cB \log B$. But X is fibered by cubic surfaces X_t with $t \in \mathbb{P}^1$, and the restriction of $-K_X$ to X_t is $-K_{X_t}$, so these cubic surfaces are expected to have $B(\log B)^{r(t)}$ points, and frequently one finds that $r(t) \geq 2$. Thus $N(U, -K_X, B) \gg cB(\log B)^2$ for every open set U.

In all known examples, the geometrically defined Nevanlinna invariant $\alpha(D)$ gives a geometric interpretation to the arithmetically defined $\beta(U/k, D)$, which is the first pole of the height zeta function $Z(U/k, D; s)$. A geometric way of computing the order of that pole is still an open question. It is easy to see that $\beta(U/k, D)$ depends only on the linear equivalence class of D (see Exercise F.17). For Fano varieties, the Néron–Severi and Picard groups are the same. In general it is tempting to try to relate $\beta(U/k, D)$ to the way in which D interacts geometrically with all of the algebraic cycles on X. This question is thoroughly discussed in the volume of Peyre [2], which contains several insights and new results.

Remark F.5.4.11. We conclude this section (and this volume) by briefly describing a few additional ways in which mathematicians study the distribution of rational points on varieties.
(a) Barry Mazur [2, 3] has proposed studying the set of rational points by comparing how it sits in X relative to the Zariski and the real topologies.

For example, assuming that $X(\mathbb{Q})$ is Zariski in X, Mazur asks whether the real closure of $X(\mathbb{Q})$ in $X(\mathbb{R})$ is always a semialgebraic set. (A *semialgebraic set* is a subset of $X(\mathbb{R})$ defined by a finite number of polynomial equations $f(x) = 0$ and a finite number of polynomial inequalities $f(x) \geq 0$.) For example, if X is an abelian variety, then the closure of $X(\mathbb{Q})$ in $X(\mathbb{R})$ will be a finite union of connected components of $X(\mathbb{R})$. However, Colliot-Thélène, Swinnerton-Dyer, and Skorobogatov [1, Example 5.2] have given a negative answer to Mazur's question. They show that if X is a smooth projective model of the surface

$$y^2 - x(4t^4 + t^2 - 4) = z^2 - (4t^4 + t^2 - 4)(x^2 + 4x - 1),$$

then $X(\mathbb{Q})$ is Zariski dense in X, $X(\mathbb{R})$ has two connected components C_1 and C_2, the set $C_1 \cap X(\mathbb{Q})$ is dense in C_1, but the real closure of $C_2 \cap X(\mathbb{Q})$ is a union of curves and points in C_2.

(b) Suppose that $X(k)$ is dense in $X(k_v)$ for some completion k_v of k. For example, we might take $k = \mathbb{Q}$ and $k_v = \mathbb{R}$. Then we can embed $X(k)$ in $X(k_v)$, choose a v-adic distance function $d_v : X(k_v) \to [0, \infty)$, and study Diophantine approximation properties of $X(k)$ within $X(k_v)$. (For a definition and properties of v-adic distance functions, see Silverman [10].) For example, one may take a point $P \in X(k_v)$ and a (decreasing) function f and ask whether the set

$$\{Q \in X(k) \mid d_v(P,Q) \leq f(H(Q))\}$$

is finite or infinite. As a specific example, suppose that $X/\mathbb{Q}$ is a projective variety of dimension n satisfying $N(X/\mathbb{Q}, D, B) \ll B^a$. Then it is not hard to show (Exercise F.12) that there exists a point $Q \in X(\mathbb{R})$ with the property that

$$d(P,Q) \gg H(P)^{-a/n} \quad \text{for all } P \in X(\mathbb{Q}).$$

Thus Q cannot be closely approximated by rational points.

(c) As already discussed in Section B.6, a very coarse measure of the distribution of rational points is given by the growth rate of the function

$$\log\log N(X/k, D, B),$$

where D is any ample divisor. In all known examples, this quantity satisfies one of the following three conditions:

$$\log\log N(X/k,D,B) \begin{cases} \sim \log\log B & \text{as } B \to \infty, \\ \sim \log\log\log B & \text{as } B \to \infty, \\ \text{is bounded} & \text{as } B \to \infty. \end{cases}$$

One might ask whether these are the only possible growth rates. In particular, we leave the reader with the intriguing question of whether

$$N(X/k, D, B) \ll \log\log B \quad \text{implies that } X(k) \text{ is finite?}$$

EXERCISES

F.1. Let X be a subvariety of an abelian variety A, let $\mathcal{L}$ be an ample line bundle on A, and define a line bundle $\mathcal{L}(-\varepsilon, s)$ on X^m as in the original proof of Faltings' theorem (see also Remark F.1.1.2). Suppose that X contains a translate of a nontrivial abelian subvariety $b + B \subset X$. Prove that $\mathcal{L}(-\varepsilon, s)$ is not ample on X^m. (*Hint.* Show that $\mathcal{L}(-\varepsilon, s)^{-1}$ is ample on $B(s) := \{(b_1, \dots, b_m) \in B^m \mid s_i b_i - s_{i+1} b_{i+1} = 0\}$.)

F.2. Let X be a subvariety of an abelian variety A whose stabilizer is trivial (but note that X may contain a translate of an abelian subvariety). Let $\mathcal{L}$ be an ample line bundle on A, and let $\mathcal{L}(-\varepsilon, s)$ be the line bundle on X^m used in the proof of Faltings' theorem (see Exercise F.1 and Remark F.1.1.2). In this exercise you will use the generalized Riemann–Roch theorem to prove that there is a constant $c > 0$ such that for all sufficiently large integers m and all sufficiently small $\varepsilon > 0$,

$$h^0\left(X^m, \mathcal{L}(-\varepsilon, s)^{\otimes d}\right) \ge c d^{m \dim X} \prod_{i=1}^{m} s_i^{2 \dim X}. \qquad (*)$$

Let n be the dimension of X. For line bundles $\mathcal{L}_1, \dots, \mathcal{L}_n$ on X, we denote their intersection number by $(\mathcal{L}_1 \cdots \mathcal{L}_n)$, and we denote the self-intersection of a line bundle $\mathcal{L}$ by $(\mathcal{L}^n)$. As usual, we let $h^0(X, \mathcal{L})$ denote the dimension of the space of sections of $\mathcal{L}$. For example, if $\mathcal{L}$ is ample and d is sufficiently large, then one knows that

$$h^0(X, \mathcal{L}^{\otimes d}) = \frac{d^n}{n!} (\mathcal{L}^n)(1 + o(1)).$$

(a) Let $\mathcal{L}$ be a line bundle on a variety Y, and let $H \subset Y$ be a hypersurface. Prove that there is an exact sequence

$$0 \to \Gamma(Y, \mathcal{L} \otimes \mathcal{O}(-H)) \to \Gamma(Y, \mathcal{L}) \to \Gamma(H, \mathcal{L}_{|H}),$$

and hence that $h^0(Y, \mathcal{L} \otimes \mathcal{O}(-H)) \ge h^0(Y, \mathcal{L}) - h^0(H, \mathcal{L}_{|H})$.

(b) Fix $\delta > 0$. Prove that $\mathcal{L}(\delta, s)$ is ample, and hence that

$$h^0(X^m, \mathcal{L}(\delta, s)^{\otimes d}) = d^{mn} \frac{(\mathcal{L}(\delta, s)^{mn})}{(mn)!}(1 + o(1)).$$

(c) Let $p_i : X^m \to X$ be the ith projection, and let H_i be a smooth hypersurface representing $p_i^* \mathcal{L}$. Prove that

$$h^0(H_i, \mathcal{L}(\delta, s)^{\otimes d}) = d^{mn-1} \frac{(p_i^* \mathcal{L} \cdot \mathcal{L}(\delta, s)^{mn-1})}{(mn-1)!}(1 + o(1))$$

and that

$$h^0(X^m, \mathcal{L}(-\varepsilon, s)^{\otimes d}) \ge h^0(X^m, \mathcal{L}(\delta, s)^{\otimes d}) - d(\delta + \varepsilon) \sum_i s_i^2 h^0(H_i, \mathcal{L}(\delta, s)^{\otimes d}_{|H_i}).$$

(d) Show that if m is sufficiently large and if $s_i \gg s_{i+1}$, then the map

$$\begin{aligned} X^m &\xrightarrow{\alpha} A^{m-1}, \\ (x_1, \dots, x_m) &\xmapsto{\alpha} (s_1x_1 - s_2x_2, \dots, s_{m-1}x_{m-1} - s_mx_m), \end{aligned}$$

is generically finite. Deduce that the sheaf $\mathcal{M} = (p_1^*\mathcal{L} \otimes \cdots \otimes p_m^*\mathcal{L})_{|\alpha(X^m)}$ satisfies $(\mathcal{L}(0,s)^{mn}) = \deg(\alpha)\,(\mathcal{M}^{mn}) > 0$.

(e) Again fix $\delta > 0$. Prove that the self-intersection number $(\mathcal{L}(\delta,s)^{mn})$ is proportional to $\prod_{i=1}^m s_i^{2\dim X}$. Using the previous part, prove that it is equal to $c(\delta)\prod_{i=1}^m s_i^{2\dim X}$ for some constant $c(\delta) > 0$.

(f) Combining (e) and (c), show that if $\varepsilon > 0$ is sufficiently small (depending on X and m), then the desired inequality $(*)$ is true.

F.3. Let $P(X)$ and $Q(X)$ be polynomials of degree $\deg(P) = 2g_1 + 1$ and $\deg(Q) = 2g_2 + 1 + \varepsilon$ respectively, where $\varepsilon \in \{0,1\}$, and assume that $P(X)Q(X)$ has no double roots. Let C be the smooth projective curve birational to the affine curve $y^2 - P(x) = z^2 - Q(x) = 0$. Prove that C has genus $2(g_1 + g_2) + \varepsilon$, and show that $W_{2g_1}(C)$ contains an abelian variety of dimension g_1. (*Hint.* Consider the map from C to the curve with affine equation $y^2 = P(x)$.)

F.4. (a) Let $E/\mathbb{Q}$ be an elliptic curve, let Δ_E and N_E be respectively the minimal discriminant and conductor of $E/\mathbb{Q}$, and write $1728\Delta_E = c_4^3 - c_6^2$ as usual. (See, e.g., Silverman [1, §III.1].) Apply the *abc* conjecture (F.3.1) to this equality (suitably divided by a gcd) to prove that

$$\max\{|\Delta_E|, |c_4^3|, |c_6^2|\} \le C_\varepsilon N_E^{6+\varepsilon}.$$

Deduce that the *abc* conjecture implies Szpiro's conjecture (F.3.2(a)) and Frey's conjecture (F.3.2(b)).

(b) Let a, b, and c be coprime integers satisfying

$$a + b + c = 0 \qquad \text{and} \qquad 2^4 \text{ divides } abc.$$

Consider the elliptic curve

$$E_{a,b,c} : y^2 = x(x-a)(x+b).$$

Prove that $\Delta_{E_{a,b,c}} = (2^{-4}abc)^2$ and $j(E_{a,b,c}) = 2^8(a^2+ab+b^2)/(abc)^2$.

(c) Prove that Frey's conjecture (F.3.2(b)) implies that the *abc* conjecture (F.3.1) is true. (*Hint.* Apply Frey's conjecture to the curve $E_{a,b,c}$ in (b).)

(d) Consider the elliptic the curve

$$E'_{a,b,c} : y^2 = x^3 - 2(a-b)x^2 + (a+b)^2x.$$

Prove that $E'_{a,b,c}$ has discriminant 2^8abc^4. Verify that the map

$$E_{a,b,c} \longrightarrow E'_{a,b,c}, \quad (x,y) \longmapsto (y^2/x^2, -y(ab+x^2)/x^2),$$

is an isogeny of degree 2. Use these facts to show that Szpiro's conjecture (F.3.2(a)) implies the *abc* conjecture with the weaker exponent $\frac{6}{5} + \varepsilon$.

F.5. Let $E/\mathbb{Q}$ be an elliptic curve and let c_4, c_6, j_E, and Δ_E be the usual quantities associated to a minimal Weierstrass equation (Silverman [1, §III.3]). Fix a complex analytic isomorphism $E(\mathbb{C}) \cong \mathbb{C}/(\mathbb{Z}+\tau\mathbb{Z})$ with $\mathrm{Im}(\tau) > 0$, and let $q = \exp(2i\pi\tau)$ and $\Delta(\tau) = (2\pi)^{-12} q \prod_{n=1}^{\infty}(1-q^n)^{24}$.

(a) Prove that the Faltings height of $E/\mathbb{Q}$ is given by the formula

$$h_{\mathrm{Falt}}(E) = \frac{1}{12}\left(\log \Delta_E - \log\left|\Delta(\tau)\,\mathrm{Im}(\tau)^6\right|\right).$$

(b) Prove that there are constants $C_0, C_1 > 0$ such that for all semistable elliptic curves $E/\mathbb{Q}$,

$$\left|\frac{1}{12}h(j_E) - h_{\mathrm{Falt}}(E)\right| \le C_0 \log(1 + h(j_E)) + C_1.$$

(c) Let $h_{\mathrm{naive}}(E) = \frac{1}{12}\log\max(|c_4^3|, |c_6^2|)$. Prove that for all $\varepsilon > 0$ there is a constant C_ε such that for all elliptic curves $E/\mathbb{Q}$,

$$|h_{\mathrm{Falt}}(E) - h_{\mathrm{naive}}(E)| \le \varepsilon h_{\mathrm{Falt}}(E) + C_\varepsilon.$$

F.6. Let $A/\mathbb{Q}$ be an abelian variety of dimension g, and let $\mathcal{A}/\mathbb{Z}$ be a Néron model for $A/\mathbb{Q}$. Prove (or take as known) that there exists a differential g-form η that generates $\Omega^g_{\mathcal{A}/\operatorname{Spec}(\mathbb{Z})}$. (The form η is called a *Néron differential* for $A/\mathbb{Q}$.) Prove that

$$\exp(h_{\mathrm{Falt}}(A/\mathbb{Q})) = \left(\frac{1}{(2\pi)^g}\int_{A(\mathbb{C})} |\eta \wedge \bar{\eta}|\right)^{-1/2}.$$

F.7. Let X be a smooth surface of bidegree $(d, 3)$ in $\mathbb{P}^1 \times \mathbb{P}^2$.

(a) If $d = 1$, show that X is a rational elliptic surface with $\kappa(X) = -1$.

(b) If $d = 2$, show that X is an elliptic K3 surface with $\kappa(X) = 0$.

(c) If $d \ge 3$, show that X is an elliptic surface with $\kappa(X) = 1$. Further, show that for all $m \ge 1$, its image $\Phi_{mK_X}(X)$ under the pluricanonical map is a rational curve C with $\kappa(C) = -1 < \kappa(X)$.

F.8. Let X be a Kodaira–Parshin surface, that is, a surface fibered over a curve of genus at least 2 such that all fibers have genus at least 2 Prove that $\kappa(X) = 2$. (*Hint.* Eliminate all of the other possibilities, namely, show that X is neither rational nor ruled, neither elliptic nor abelian, neither a K3 surface nor an Enriques surface.)

F.9. Let $X = \mathbb{P}^2$, and let D_1, D_2, D_3 be the divisors defined respectively by $x = 0$, $y = 0$ and $(x-y)z - (x+y)^2 = 0$. Set $D = D_1 + D_2 + D_3$. Clearly, D is effective and reduced.

(a) Prove that D does not have normal crossings (at the point $(0, 0, 1)$).

(b) For any unit $\varepsilon \in R_k^*$ and any $i, j \in \mathbb{Z}$, let

$$P_{i,j}^{\varepsilon} = (\varepsilon^i, 1, \varepsilon^i + 3 - 4(\varepsilon^{ij} - 1)/(\varepsilon^i - 1)) \in \mathbb{P}^2(k).$$

Prove that $P_{i,j}^{\varepsilon}$ is integral with respect to the divisor D. Conclude that Vojta's conjecture (F.5.3.2), and more specifically (F.5.3.6), is false for D. Thus the normal crossings requirement in Vojta's conjecture is needed.

F.10. (a) Let X be a smooth projective variety, and let D be an effective normal crossings divisor. Use Vojta's conjecture (F.5.3.2) to prove that outside of a proper Zariski closed subset, all S-integral points with respect to D satisfy

$$h_D(P) \le \epsilon h_E(P) + h_{-K_X}(P) + C_\epsilon.$$

(b) Let X be a projective variety such that $mK_X = 0$ for some $m \ge 1$. For example, X could be a K3 or an Enriques surface. Let D be an ample effective divisor on X. Show that Vojta's conjecture (F.5.3.2) implies that the set of S-integral points with respect to D is not Zariski dense in X.
(c) Repeat (b) when X is a cubic surface in $\mathbb{P}^3$ and D is the union of two hyperplane sections.

F.11. Prove that the generalized Vojta conjecture (F.5.3.8) implies the *abc* conjecture (F.3.1). (*Hint.* If $a+b+c=0$, then the point $(a^{1/N}, b^{1/N}, c^{1/N})$ is an algebraic point on the Fermat curve $X^N+Y^N+Z^N=0$. Apply Vojta's conjecture to this curve.)

F.12. Let $X \subset \mathbb{P}^m$ be a smooth projective variety of dimension n defined over $\mathbb{Q}$, and assume that there is a constant $a>0$ such that $N(X(\mathbb{Q}), H, B) \ll B^a$. In particular, $X(\mathbb{Q})$ has a large number of rational points, so we can study approximation of real points by rational points.
(a) Show that $X(\mathbb{R})$ has dimension n as a real differentiable manifold. Take any natural distance function on $\mathbb{P}^m(\mathbb{R})$ and use it to induce a distance function $d(x,y)$ on $X(\mathbb{R})$. For any natural volume function on $X(\mathbb{R})$, prove that the volume of a "ball"

$$\mathcal{B}(x,r) = \{y \in X(\mathbb{R}) \mid d(x,y) \le r\}$$

satisfies $\operatorname{Vol}(\mathcal{B}(x,r)) \gg\ll r^n$.
(b) Construct a point $x \in X(\mathbb{R})$ such that

$$d(x,y) \gg H(y)^{-a/n} \qquad \text{for all } x \in X(\mathbb{Q}).$$

(*Hint.* This is trivial if $X(\mathbb{Q})$ is not dense in $X(\mathbb{R})$. Otherwise, consider the sets $\cup_{H(y)\le B}\mathcal{B}(y, cB^{-a/n})$ and use the fact that an (infinite) intersection of a decreasing sequence of compact sets is not empty.)

F.13. Let D, D_1, D_2 be ample divisors on X, and let U be an open subset of X. In this exercise we compare the formal properties of the Nevanlinna invariant $\alpha(D)$ and the abscissa of convergence $\beta(U/k, D)$ of the height zeta function $Z(U/k, D; s)$. Since U/k is fixed, we will ease notation by writing $\beta(D)$ for $\beta(U/k, D)$.
(a) Prove that $\alpha(mD) = \alpha(D)/m$ and $\beta(mD) = \beta(D)/m$.
(b) If $D_1 \ge D_2$, prove that $\alpha(D_1) \le \alpha(D_2)$. If further

$$U \cap \operatorname{supp}(D_1 - D_2) = \emptyset,$$

prove that $\beta(D_1) \le \beta(D_2)$. Is this true without the condition on U?
(c) Prove the inequalities

$$\alpha(D_1 + D_2) \le \frac{\max(\alpha(D_1), \alpha(D_2))}{2}, \quad \beta(D_1 + D_2) \le \frac{\max(\beta(D_1), \beta(D_2))}{2}.$$

Prove the slightly stronger estimate $\alpha(D_1)^{-1} + \alpha(D_2)^{-1} \le \alpha(D_1 + D_2)^{-1}$. (*Hint.* Use the characterization $\alpha(D)^{-1}K_X + D \in \partial\mathrm{NS}_{\mathrm{eff}}(X)$.) Is this stronger inequality also true for β_U?

F.14. Let $0 \le m \le n-2$, and let L_0 be the linear subspace of $\mathbb{P}^n$ defined by the equations $x_{m+1} = \cdots = x_n = 0$, so $L_0 \cong \mathbb{P}^m$. Let X be the blowup of $\mathbb{P}^n$ along L_0. Explicitly, X is the variety

$$X = \{((x_0,\dots,x_n),(y_{m+1},\dots,y_n)) \in \mathbb{P}^n \times \mathbb{P}^{n-m-1} \mid x_iy_j - x_jy_i = 0\},$$

where the equations are taken for all $m+1 \le i,j \le n$. Let $\pi : X \to \mathbb{P}^n$ be the natural projection, let $L \in \mathrm{Div}(X)$ be the pullback of a hyperplane by π, and let $E = L_0 \times \mathbb{P}^{n-m-1} \in \mathrm{Div}(X)$ be the exceptional divisor of the blowup.

(a) Show that π is an isomorphism from $U = X \smallsetminus E$ to $\mathbb{P}^n \smallsetminus L_0$. Prove that $\mathrm{Pic}(X)$ is a free group of rank 2, given explicitly by $\mathrm{Pic}(X) = \mathbb{Z}[L] \oplus \mathbb{Z}[E]$. Prove that $K_X = -(n+1)L + (n-m-1)E$, and that a divisor class $D = aL - bE$ is ample if and only if $a > b > 0$.

(b) Let $a > b > 0$. Prove that

$$\alpha(aL - bE) = \min\left\{\frac{n+1}{a}, \frac{m+2}{a-b}\right\}.$$

(c) Let $P = ((x_0,\dots,x_n),(y_{m+1},\dots,y_n)) \in X(\mathbb{Q})$ with $x_i, y_j \in \mathbb{Z}$ and $\gcd(x_0,\dots,x_n) = \gcd(y_{m+1}\dots,y_n) = 1$. Prove that

$$H_L(P) = \max_i |x_i| \quad \text{and} \quad H_{2L-E}(P) = \max_i |x_i| \cdot \max_j |y_j|.$$

Deduce from this the formula $H_E(P) = \max_i |x_i| / \max_j |y_j|$. (N.B. Since multiplicative Weil height functions are defined only up to constant multiples, what you are proving is that the given functions are particular Weil height functions for the specified divisors.)

(d) Prove that the distribution of the rational points on the exceptional divisor E is given by

$$N(E/\mathbb{Q}, aL - bE, B) \gg\ll \begin{cases} B^{(m+1)/(a-b)} & \text{if } \frac{m+1}{a-b} > \frac{n-m}{b}, \\ B^{(m+1)/(a-b)} \log B & \text{if } \frac{m+1}{a-b} = \frac{n-m}{b}, \\ B^{(n-m)/b} & \text{if } \frac{m+1}{a-b} < \frac{n-m}{b}. \end{cases}$$

(*Hint.* Show that the left-hand side is $N((\mathbb{P}^m \times \mathbb{P}^{n-m-1})/\mathbb{Q}, \mathcal{O}(a-b,b), B)$ and use Exercise F.15 below.)

(e) Let $P = ((x_0,\dots,x_n),(y_{m+1},\dots,y_n)) \in U(\mathbb{Q})$ be a point with integer coordinates, and let $d = \gcd(x_{m+1},\dots,x_n)$. Set

$$N = \max_{0 \le i \le m} |x_i| \quad \text{and} \quad M = \max_{m+1 \le j \le n} |y_j| = \max_{m+1 \le j \le n} |x_j|/d.$$

Prove that $H_{aL-bE}(P) = \max(N, dM)^{a-b} M^b$. Use this formula to prove that

$$N(U/\mathbb{Q}, aL - bE, B) \gg\ll \begin{cases} B^{(m+2)/(a-b)} & \text{if } \frac{n-m-1}{n+1} < \frac{b}{a}, \\ B^{(n+1)/a} \log B & \text{if } \frac{n-m-1}{n+1} = \frac{b}{a}, \\ B^{(n+1)/a} & \text{if } \frac{n-m-1}{n+1} > \frac{b}{a}. \end{cases}$$

When is $N(U/\mathbb{Q}, aL - bE, B)$ greater than, less than, or comparable to $N(E/\mathbb{Q}, aL - bE, B)$?

F.15. (a) Let $S_1,\dots,S_m$ be sets, and let $H_i : S_i \to [1,\infty)$ be functions that can be used for counting as in Lemma F.5.4.5. Suppose that

$$N(S_i,H_i,B) := \#\{x \in S_i \mid H_i(x) \le B\} \sim c_i B^{a_i} (\log B)^{b_i}$$

for each $1 \le i \le m$. Define a function on the product by

$$\begin{aligned} H : S = S_1 \times \cdots \times S_m &\longrightarrow [1,\infty), \\ H(x_1,\dots,x_m) &= H_1(x_1)\cdots H_m(x_m). \end{aligned}$$

Prove that the counting function for the product,

$$N(S,H,B) := \#\{x \in S \mid H(x) \le B\},$$

satisfies

$$N(S,H,B) \sim cB^a (\log B)^b$$
$$\text{with } a = \max_i a_i \text{ and } b = -1 + \sum_{i \text{ with } a_i = a} (b_i + 1).$$

(b) Let $X = \mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_m}$, and let $\mathcal{L}$ be the line bundle $\mathcal{O}(d_1,\dots,d_m)$ on X with $d_1,\dots,d_m$ positive integers. Apply (a) to determine the asymptotic behavior of $N(X,\mathcal{L},B)$.

F.16. Let $F_1,\dots,F_r \in \mathbb{Z}[X_0,\dots,X_n]$ be homogeneous polynomials, and define

$$\begin{aligned} X(\mathbb{Z}) &= \{x \in \mathbb{Z}^{n+1} \mid F_1(x) = \cdots = F_r(x)\}, \\ M(B) &= \{x \in X(\mathbb{Z}) \mid \max |x_i| \le B\}, \\ N(B) &= \{x \in X(\mathbb{Z}) \mid \gcd(x) = 1 \text{ and } \max |x_i| \le B\}. \end{aligned}$$

Prove that

$$M(B) = \sum_{d \le B} N(B/d) \quad \text{and} \quad N(B) = \sum_{d \le B} \mu(d) M(B/d),$$

where μ is the Möbius function.

F.17. (a) Let S be a set, and let $H, H' : S \to \mathbb{R}$ be two functions used for counting as in Lemma F.5.4.5. Suppose that $C_1 H(x) \le H'(x) \le C_2 H(x)$ for all $x \in S$. Prove that the corresponding counting functions satisfy

$$N(S,H,C_1B) \le N(S,H',B) \le N(S,H,C_2B).$$

(b) Let D and D' be linearly equivalent ample divisors on a variety X. Use (a) to prove that

$$N(X/k,D,B) \gg\ll B^a (\log B)^b \quad \text{iff} \quad N(X/k,D',B) \gg\ll B^a (\log B)^b.$$

In particular, up to $\gg\ll$, the growth of the counting function $N(X/k,D,B)$ depends only the equivalence class of D and is independent of the choice of a particular Weil height H_D for D.

(c) In a similar vein, prove that the abscissa of convergence of the height zeta function $Z(X/k,D,s)$ depends only on the class of D in $\mathrm{NS}(X)$, i.e., it depends only on the algebraic equivalence class of D.

F.18. This exercise shows that one may restrict to normal varieties when studying counting functions on varieties. Let X be a singular variety and let $\nu : X' \to X$ be its normalization. Let $\mathcal{L}$ be an ample line bundle on X, so $\mathcal{L}' = \nu^*\mathcal{L}$ is ample on X'. Let Z be the locus of nonnormal points on X and $Z' = \nu^{-1}(Z)$. Set $U = X \setminus Z$ and $U' = X' \setminus Z'$. Prove that

$$\#\{y \in U(k) \mid H_{\mathcal{L}}(y) \le B\} = \#\{x' \in U'(k) \mid H_{\mathcal{L}'}(x) \le B\}.$$

F.19. Try to prove or disprove some of the conjectures described in this book. (Unfortunately, the authors do not know how to solve this exercise!)

References

Abbès, A., Ullmo, E.

[**2**] A propos de la conjecture de Manin pour les courbes elliptiques modulaires. *Compo. Math.* **103** (1996), 269–286.

Abramovic, D.

[1] A high fibered power of a family of varieties of general type dominates a variety of general type. *Invent. Math.* **128** (1997), 481–494.

[**2**] Uniformity of stably integral points on elliptic curves. *Invent. Math.* **127** (1997), 307–317.

Abramovic, D., Harris, J.

[1] Abelian varieties and curves in $W_d(C)$. *Compositio Math.* **78** (1991), 227–238.

Abramovic, D., Voloch, J.F.

[1] Towards a proof of the Mordell–Lang conjecture in characteristic *p*. *International Math. Research Notices* **2** (1992), 103–115.

[**2**] Lang's conjecture, fibered powers and uniformity. *New York J. Math.* **II** (1996), 20–34.

Apostol, T.

[1] *Introduction to Analytic Number Theory,* Springer-Verlag, 1976.

Arbarelo, E., Cornalba, M., Griffiths, P.A., Harris, J.

[1] *Geometry of Algebraic Curves,* Springer-Verlag, 1985.

Artin, M.

[1] Néron models. In *Arithmetic Geometry,* Cornell, Silverman, eds., Springer-Verlag, 1986, 213–230.

Artin, M., Winters, G.

[1] Degenerate fibres and stable reduction of curves. *Topology* **10** (1971), 373–383.

Atiyah, M.F., Macdonald, I.G.

[1] *Introduction to Commutative Algebra,* Addison-Wesley, 1969.

Atiyah, M.F., Wall, C.

[1] Cohomology of groups. In *Algebraic Number Theory,* Cassels, Fröhlich, eds., Academic Press, 1967, 94–115.

Baker, A.

[1] *Transcendental Number Theory,* Cambridge University Press, 1975.

Batyrev, V., Manin, Y.

[1] Sur le nombre des points rationnels de hauteur borné des variétés algébriques. *Math. Ann.* **286** (1990), 27–44.

Batyrev, V., Tschinkel, Y.

[1] Rational points on some Fano cubic bundles. *CRAS* **323** (1996), 41–46.

[2] Manin's conjecture for toric varieties. *J. of Alg. Geometry* **7** (1998), 15–53.

Belyi, G.

[1] Galois extensions of a maximal cyclotomic field. *Izv. Akad SSSR ser. Mat.* **43** (1979), 267–276.

Billard, H.

[1] Propriétés arithmétique d'une surface K3. *Compositio Math.* **108** (1997), 247–275.

[2] Sur la répartition des points sur les surfaces elliptiques. *J. reine angew. Math* **505** (1998), 45–71.

[3] Répartition des points rationnels des surfaces géométriquement réglées rationnelles. *Astérisque* **251** (1998), 79–89.

Birch, B.

[1] Forms in many variables. *Proc. Royal Soc. of London* **265 A** (1962), 245–263.

Bogomolov, F.

[1] Points of finite order on an abelian variety. *Math. USSR Izv.* **17** (1981).

Bombieri, E.

[1] The Mordell conjecture revisited. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **17** (1990), 615–640.

Bombieri, E., van der Poorten, A.J.

[1] Some quantitative results related to Roth's theorem. *J. Austral. Math. Soc.* **45** (198), 233–248.

Bombieri, E., Zannier, U.

[1] Algebraic points on subvarieties of $\mathbb{G}_m^n$. *Internat. Math. Res. Notices* **7** (1995), 333–347.

[2] Heights of algebraic points on subvarieties of abelian varieties. preprint.

Bost, J.-B.

[1] Introduction to compact Riemann surfaces, jacobians and abelian varieties. In *From Number Theory to Physics* (Colloque Les Houches), Springer-Verlag, 1992.

[2] Intrinsic heights of stable varieties and abelian varieties. *Duke Math. J.* **82** (1995), 21–70.

Bost, J.-B., David, S.
[1] Hauteurs theta. preprint.
Bost, J.-B., Gillet H., Soulé, C.
[1] Heights of projective varieties and positive Green forms. *J. American Math. Soc.* **7** (1994), 903–1022.
Bosch, Lütkebohmert, Raynaud, M.
[1] *Néron Models,* Springer-Verlag, 1990.
Bourbaki, N.
[1] *Groupes et Algèbres de Lie,* Masson, Paris, 1972.
Bouscaren, E.
[1] *Model Theory and Algebraic Geometry. An introduction to E. Hrushovski's proof of the geometric Mordell-Lang conjecture,* Lect. Notes in Math. 1696, Springer-Verlag, 1998.
Breuil, C., Conrad, B., Diamond, F., Taylor, R.
[1] Every elliptic curve over $\mathbb{Q}$ is modular. announcement, July 1999.
Brumer, A., Kramer, K.
[1] The conductor of an abelian variety. *Compositio Math.* **92** (1994), 227–248.
Buium, A.
[1] Intersections in jet spaces and a conjecture of Serge Lang. *Annals of Math.* **136** (1992), 583–593.
[2] Effective bounds for the geometric Lang conjecture. *Duke J. Math.* **71** (1993), 475–499.
[3] On a question of Mazur. *Duke Math. J.* **75** (1994), 639–644.
Burnol, J.-F.
[1] Weierstrass points on arithmetic surfaces. *Invent. Math.* **107** (1992), 421–432.
Call, G.
[1] Counting geometric points on families of abelian varieties. *Math. Nachrichten* **166** (1994), 167–192.
Call, G., Silverman, J.
[1] Canonical heights on varieties with morphisms. *Compositio Math.* **89** (1993), 163–205.
Caporaso, L., Harris, J., Mazur, B.
[1] Uniformity of rational points. *J. American Math. Soc.* **10** (1997), 1–35.
Cassels, J.W.S.
[1] Diophantine equations with special reference to elliptic curves. *J. London Math. Soc.* **41** (1966), 193–291.
[2] *Lectures on Elliptic Curves,* Oxford University Press, 1990.

Chabauty, C.

[1] Sur les points rationnels des variétés algébriques dont l'irrégularité est supérieure à la dimension. *C.R. Acad. Sci. Paris* **212** (1941), 1022–1024.

Chambert-Loir, A.

[1] Points de petites hauteurs sur les variétés semi-abéliennes. *Ann. Ecole Norm. Sup.* to appear.

Chandrasekharan, K., Narasimhan, R.

[1] Functional equations with multiple Gamma factors and the average of arithmetical functions. *Annals of Math.* **76** (1962), 93–136.

Chinburg, T.

[1] An introduction to Arakelov intersection theory. In *Arithmetic Geometry,* Cornell, Silverman, eds., Springer-Verlag, 1986, 289–308.

Coleman, R.

[1] p-adic integrals and torsion points on curves. *Annals of Math.* **121** (1985), 111–168.

[2] Effective Chabauty. *Duke J.* **52** (1985), 765–770.

Colliot-Thélène, J.-L., Skorobogatov, A., Swinnerton-Dyer, P.

[1] Double fibres and double covers, paucity of rational points. *Acta Arithmetica* **79** (1997), 113–135.

Cornell, G., Silverman, J.H. (eds.)

[1] *Arithmetic Geometry,* Springer-Verlag, 1986.

Davenport, H., Roth. K.F.

[1] Rational approximations to algebraic numbers. *Mathematika* **2** (1955), 1–20.

David, S.

[1] Minoration de hauteurs sur les variétés abéliennes. *Bull. Soc. Math. France* **121** (1993), 509–544.

David, S., Hindry, M.

[1] Minoration de la hauteur de Néron–Tate sur les variétés abéliennes de type CM. *J. Reine Angew. Math.* to appear.

David, S., Philippon, P.

[1] Minorations des hauteurs normalisées des sous-variétés de variétés abéliennes . In *International Conference on Discrete Mathematics and Number Theory (Tiruchirapelli, 1996),* K. Murty and M. Waldschmidt , eds., Contemp. Math., 1998, 3–17 .

Davis, M., Matyasevic, Y., Putnam, H., Robinson, J.

[1] Hilbert's tenth problem, Diophantine equations: positive aspects of a negative solution. In *Symp. Pure Math. 28,* AMS, 1976, 323–378.

Debarre, O.

[1] Variétés de Fano. *Séminaire Bourbaki* exposé 827, mars 1997.

Debarre, O,. Fahlaoui, R.

[1] Abelian varieties and curves in $W_d^r(C)$ and points of bounded degree on algebraic curves. *Compositio Math.* **88** (1993), 235–249.

Debarre, O., Klassen, M.
[1] Points of low degree on smooth plane curves. *J. Reine Angew. Math.* **446** (1994), 81–87.
Demjanenko, V.A.
[1] Rational points of a class of algebraic curves. *Izv. Akad. Nauk SSSR Ser. Mat.* **30** (1966), 1373–1396. transl. in *AMS Translations* **66** (1968), 246–272.
de Diego, T.
[1] Points rationnels sur les familles de courbes de genre au moins 2. *J. of Number Theory* **67** (1997), 85–114.
Diamond, F.
[1] On deformation rings and Hecke rings. *Annals of Math.* **144** (1996), 137–166.
Edixhoven, B., Evertse, J.-H.
[1] *Diophantine approximation and abelian varieties,* Lect. Notes in Math. 1566, Springer-Verlag, 1993.
Elkies, N.
[1] A, B, C implies Mordell. *International Math. Res. Notices* **7** (1991), 99–109.
Evertse, J.-H.
[1] On equations in S-units and the Thue–Mahler equation. *Invent. Math.* **75** (1984), 561–584.
[2] An explicit version of Faltings' product theorem and an improvement of Roth's lemma. *Acta Arithmetica* **73** (1995), 215–248.
Fadeev, D.
[1] On the divisor class group of some algebraic curves. *Doklady* **136** (1961), 296–298. *Sov. Mat.* **2** (1961), 67–69.
Faltings, G.
[1] Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. *Invent. Math.* **73** (1983), 349–366.
[2] Diophantine approximation on abelian varieties. *Annals of Math.* **133** (1991), 549–576.
[3] The general case of Lang's conjecture. In *Symposium in Algebraic geometry,* Barsotti, eds., Acad. Press, 1994, 175–182.
Faltings, G., Wüstholz, G.
[1] Approximations on projective spaces. *Invent. Math.* **116** (1994), 109–138.
Feldman, I.
[1] An effective refinement of the exponent in Liouville's theorem. *Izv. Akad. Nauk SSSR Mat.* **35** (1971), 973–990. *Math. USSR Izv.* **5** (1971), 985–1002.
Flynn, E.
[1] Sequences of rational torsion points on abelian varieties. *Invent. Math.* **106** (1991), 433–442.
[2] On a theorem of Coleman. *Manuscripta Math.* **88** (1995), 447–456.

Fontaine, J.-M.

[1] Il n'y a pas de variété abélienne sur $\mathbb{Z}$. *Invent. Math.* **81** (1985), 515–538.

Franke, J., Manin, Y., Tschinkel, Yu.

[1] Rational points of bounded height on Fano varieties. *Invent. Math.* **95** (1989), 421–435.

Frey, G.

[1] Links between stable elliptic curves and certain diophantine equations. *Ann. Univ. Sarav.* **1** (1986), 1–40.

[2] Elliptic curves and solutions of $A - B = C$. In *Sém. Théorie des Nombres Paris* 1985–86, C. Goldstein, eds., Birkhäuser, 1987, 39–51.

[3] Curves with infinitely many points of fixed degree. *Isr. J. Math.* **85** (1994), 79–83.

[4] A remark about isogenies of elliptic curves over quadratic fields. *Compositio Math.* **58** (1986), 133–134.

Fulton, W.

[1] *Algebraic Curves,* Benjamin, 1969.

[2] *Intersection Theory,* Springer-Verlag, 1984.

Griffiths, P., Harris, J.

[1] *Principles of Algebraic Geometry,* Wiley, 1978.

Gross, B., Rohrlich, D.

[1] Some Results on the Mordell–Weil Group of the Jacobian of the Fermat Curve. *Invent. Mat.* **44** (1978), 201–224.

Gross, R.

[1] A note on Roth's theorem. *J. Number Theory* **36** (1990), 127–132.

Gubler, W.

[1] Höhentheorie. *Math. Ann.* **298** (1994), 427–456.

Gunning, R.

[1] *Lectures on Riemann Surfaces,* Princeton U.Press, 1966.

Hardy, G.H., Wright, E.M.

[1] *An Introduction to the Theory of Numbers* (4th edition), Oxford University Press, 1960.

Harris, J., Silverman, J.H.

[1] Bielliptic curves and symmetric curves. *Proc. AMS* **112** (1991), 347–356.

Hartshorne, R.

[1] *Algebraic Geometry,* Springer-Verlag, 1977.

Heath-Brown, D.R.

[1] Cubic forms in ten variables. *Proc. London Math. Soc.* **47** (1983), 225–257.

[2] A new form of the circle method and its applications to quadratic forms. *J. Reine Angew. Math.* **481** (1996), 149–206.

Herstein, I.N.

[1] *Topics in Algebra* (2nd edition), Xerox College Publishing, 1975.

Hilton, P., Stammbach U.
[1] *A course in homological algebra,* Springer-Verlag, 1970.
Hindry, M.
[1] Autour d'une conjecture de Serge Lang. *Invent. Math.* **94** (1988), 575–603.
[2] Sur les conjectures de Mordell et Lang (d'après Vojta, Faltings et Bombieri). *Astérisque* **209** (1992), 39–56.
[3] Points quadratiques sur les courbes. *CRAS* série A **305** (1987), 219–221.
Hindry, M., Silverman, J.
[1] Canonical heights and integral points on elliptic curves. *Invent. Math.* **93** (1988), 419–450.
Hooley, C.
[1] Nonary cubic forms I, II, III. *Journal Reine Angew. Math.* **386** (1988), 32–39. **415** (1991), 95–165. **456** (1994), 53–63.
Hrushovski, E.
[1] The Mordell–Lang conjecture for function fields. *Journal AMS* **9** (1996), 667–690.
Iitaka, S.
[1] *Algebraic Geometry, an Introduction to Birational Geometry of Algebraic Varieties,* Springer-Verlag, 1982.
Kamienny, S.
[1] Torsion points on elliptic curves and q-coefficients of modular forms. *Invent. Math.* **109** (221–229), 1992.
Kenku, M.A., Momose, F.
[1] Torsion points on elliptic curves defined over quadratic fields. *Nagoya Math. J.* **109** (1988), 125–149.
Kollar, J.
[1] *Rational curves on algebraic varieties,* Ergebnisse der Math., Springer-Verlag, 1996.
Kolyvagin, V.A.
[1] Finiteness of $E(\mathbb{Q})$ and $Ш(\mathbb{Q})$ for a class of Weil curves. *Math. USSR Izv.* **32** (1989), 523–542.

Lang, S.
[1] Integral points on curves. *Publ. I.H.E.S. Math.* Paris, 1960.
[2] *Algebra* (2nd edition), Addison-Wesley, 1984.
[3] *Abelian Varieties*, Interscience Pub., 1959.
[4] *Algebraic and Abelian Functions* (2nd edition), Springer-Verlag, 1982.
[5] *Elliptic Curves: Diophantine Analysis*, Springer-Verlag, 1978.
[6] *Fundamentals of Diophantine Geometry*, Springer-Verlag, 1983.
[7] *Introduction to Arakelov Theory*, Springer-Verlag, 1988.
[8] *Number Theory III, Diophantine Geometry*, Springer-Verlag, 1991.
[9] *Algebraic Number Theory*, Springer-Verlag, 1986.
[10] Some applications of the local uniformisation theorem. *Amer. J. Math.* **76** (1954), 362–374.
[11] *Elliptic Functions*, GTM 112, Springer-Verlag, New York, 1987.
[12] Division points on curves. *Ann. Mat. Pura Appl.* **LXX** (1965), 229–234.
[13] Hyperbolic and Diophantine Analysis. *Bulletin AMS* **14** (1986), 159–205.

Lange, Birkenhake
[1] *Complex Abelian Varieties*, Springer-Verlag, 1992.

Laurent, M.
[1] Equations diophantiennes exponentielles. *Invent. Math.* **78** (1984), 299–327.

Leprévost, F.
[1] Torsion sur des familles de courbes de genre g. *Manuscripta Math.* **75** (1992), 303–326.
[2] Sur une conjecture sur les points de torsion rationnels des jacobiennes de courbes. *J. Reine Angew. Math.* **473** (1996), 59–68.

Lewis, D.J., Mahler, K.
[1] On the representation of integers by binary forms. *Acta Arith.* **6** (1961), 333–363.

Liardet, P.
[1] Sur une conjecture de Serge Lang. *Astérisque* **24–25** (1975), 187–209.

Liouville, J.
[1] Sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques. *Journal de Math.Pures et Appl.* **16** (1851), 133–149.

Lockhart, P., Rosen, M., Silverman, J.
[1] An upper bound for the conductor of an abelian variety. *J. Algebraic Geometry* **2** (1993), 569–601.

Mahler, K.

[1] Zur Approximation algebraischer Zahlen. *Math. Ann.* **107** (1933), 691–730.

[2] Über die rationalen Punkte auf Kurven vom Geschlecht Eins. *J. Reine Angew. Math.* **170** (1934), 168–178.

[3] An inequality for the discriminant of a polynomial. *Mich. Math. J.* **11** (1964), 257–262.

Manin, Y.

[1] Rational points of algebraic curves over function fields. *Izv. Akad. Nauk SSSR* **27** (1963), 1395–1440 (Russian). *Translations II AMS* **50** (1966), 189–234.

[2] *Cubic Forms* (2nd ed.), North-Holland, 1986.

[3] Notes on the arithmetic of Fano threefolds. *Compos. Math.* **85** (1993), 37–55.

[4] Cyclotomic fields and modular curves. *Russian Math. Surveys* **26** **(6)** (1971), 7–78.

[5] The p-torsion of elliptic curves is uniformly bounded. *Izv. Akad Nauk USSR* **33** (1969), 459–465. *AMS Transl. Math. USSR Izv.* **3** (1969), 433–438.

Manin, Y., Tsfasman, J.

[1] Rational varieties, algebra, geometry and arithmetic. *Uspekhi Mat. Nauk* **41** (1986), (Russian). *Russian Math. Surveys* **41** (1986), 51–116.

Manin, Y. Tschinkel, Y.

[1] Rational points of bounded height on Del Pezzo surfaces. *Compositio Math.* **85** (1993), 315–332.

Masser, D.

[1] Small values of heights on families of abelian varieties. In *Diophantine approximation and transcendance theory.* Lect. Notes in Math. 1290, G. Wüstholz, eds., Springer-Verlag, 1985, 109–148.

Masser, D., Wüstholz, G.

[1] Fields of large transcendence degree generated by values of elliptic functions. *Invent. Math.* **72** (1983), 407–464.

[2] Periods and minimal abelian subvarieties. *Annals of Math.* **137** (1993), 407–458.

[3] Isogenies estimates for abelian varieties and finiteness theorems. *Annals of Math.* **137** (1993), 459–472.

Matjasevic, Y.

[1] Diophantovost' perechislimych mnozhestv. *Doklady Akad. Nauk SSSR* **191** (1970), 279–282. Enumerable sets are diophantine. *Soviet Math. Doklady* **11** (1970), 354–358.

Matsumura, H.

[1] *Commutative Algebra,* Benjamin, 1970.

Mazur, B.

[1] Modular curves and the Eisenstein ideal. *Pub. Math. IHES* **47** (1978), 33–186.

[2] The topology of rational points. *Experimental Mathematics* **1** (1992), 35–45.

[3] Speculation about the topology of rational points. *Astérisque* **228** (1995), 165–181.

McQuillan, M.

[1] Division points on semi-abelian varieties. *Invent. Math.* **120** (1995), 143–159.

Merel, L.

[1] Borne pour la torsion des courbes elliptiques sur les corps de nombres. *Invent. Math.* **124** (1996), 437–449.

Mestre, J.-F.

[1] Formules explicites et minoration des conducteurs de variétés algébriques. *Comput. Math.* **58** (1986), 209–232.

Mignotte, M.

[1] Quelques remarques sur l'approximation rationnelle des nombres algébriques. *J. reine angew. Math.* **268/269** (1974), 341–347.

Milne, J.

[1] Abelian varieties. In *Arithmetic Geometry,* Cornell, Silverman, eds., Springer-Verlag, 1986, 103–150.

[2] Jacobian Varieties. In *Arithmetic Geometry,* Cornell, Silverman, eds., Springer-Verlag, 1986, 167–212.

Mordell, L.J.

[1] On the rational solutions of the indeterminate equations of the third and fourth degrees. *Proc. Camb. Philos. Soc.* **21** (1922), 179–192.

Moret-Bailly, L.

[1] Hauteurs et classes de Chern sur les surfaces arithmétiques. *Astérisque* **183** (1990), 37–58.

Mori, S.

[1] Projective manifolds with ample tangent bundles. *Annals of Math.* **110** (1979), 593–606.

Morita, Y., Sato, A.

[1] Distribution of rational points on hyperelliptic surfaces. *Tôhoku Math. J.* **44** (1992), 345–358.

Moriwaki, A.

[1] Remarks on rational points of varieties whose cotangent bundles are generated by global sections. *Math. Research Letters* **2** (1995), 113–118.

Mumford, D.

[1] A remark on Mordell's conjecture. *Amer. J. Math.* **87** (1965), 1007–1016.

[2] *Abelian Varieties,* Oxford U. Press, 1970.

[3] *Curves and Their Jacobians,* Michigan Univ. Press, 1975.

[4] *Algebraic Geometry I. Complex Projective Varieties,* Springer-Verlag, 1976.

[5] *Tata Lectures on Theta Functions I and II,* Progr. in Math. 28 & 43, Birkhäuser, 1984 & 1985.

[6] *The Red Book of Varieties and Schemes,* Lect. Notes in Math. 1358, Springer-Verlag, 1988.

[7] On equations defining abelian varieties. *Invent. Mat.* **1** (1966), 287–354.

Mumford, D., Fogarty, J.

[1] *Geometric Invariant Theory* Ergebnisse der Math. 34, Springer-Verlag, Berlin, 1965 and 1982.

Murty, K.

[1] Modular elliptic curves. In *Seminar on Fermat's Last Theorem,* K. Murty, eds., AMS, 1995, 1–38.

Néron, A.

[1] Modèles minimaux des variétés abéliennes sur les corps locaux et globaux. *Publ. I.H.E.S. Math.* **21** (1964), 361–482.

[2] Quasi-fonctions et hauteurs sur les variétés abéliennes. *Annals of Math.* **82** (1965), 249–331.

Nishimura

[1] Some remarks on rational points. *Mem. Coll. Sci. Kyoto* ser. A **29** (1955), 189–192.

Noguchi, J.

[1] A higher dimensional analog of Mordell's conjecture over function fields. *Math. Annalen* **258** (1981), 207–212.

Northcott, D.G.

[1] An inequality in the theory of arithmetic on algebraic varieties. *Proc. Camb. Philos. Soc.* **45** (1949), 502–509.

[2] A further inequality in the theory of arithmetic on algebraic varieties. *Proc. Camb. Philos. Soc.* **45** (1949), 510–518.

[3] Periodic points of an algebraic variety. *Annals of Math.* **51** (1950), 167–177.

Oesterlé, J.

[1] Nouvelles approches du théorème de Fermat. *Séminaire Bourbaki* **694** (1988).

Pacelli, P.L.

[1] Uniform boundedness for rational points. *Duke Math. J.* **88** (1997), 77–102.

[2] Uniform bounds for stably integral points on elliptic curves. *Proc. Amer. Math. Soc.* to appear.

Parshin, A.

[1] The Bogomolov–Miyaoka–Yau inequality for arithmetical surfaces and its applications. In *Sém. Théorie des Nomb. Paris 1986–87, Progress in Math.* **75**, Birkhäuser, 1986–87, 299–311.

Peyre, E.

[1] Hauteurs et nombres de Tamagawa sur les variétés de Fano. *Duke Math. J.* **79** (1995), 101–128.

[2] *Nombre et répartition de points de hauteur bornée,* Astérisque 251, 1998.

Philippon, P.

[1] Sur des hauteurs alternatives I, II et III. *Math. Ann.* **289** (1991), 255–283. *Ann. Inst. Fourier* **44** (1043–1065), 1994. *J. Math. Pures et Appl.* **74** (345–365), 1995.

Pila, J.

[1] Density of integral and rational points on varieties. *Astérisque* **228** (1995), 183–187.

Poizat, B.

[1] H.L.M. (Hrushovski, Lang, Mordell. *Bourbaki* exposé 811, 1996. Mordel413-425

Poonen, B.

[1] Mordell–Lang plus Bogomolov. *Inventiones Math.* **137** (1999), 413–425.

Poonen, B., Schaefer, E.

[1] Explicit descent for Jacobians of cyclic covers of the projective lines. *J. reine angew. Math.* **488** (1997), 141–188.

Raynaud, M.

[1] Courbes sur une variété abélienne et points de torsion. *Invent. Math.* **71** (1983), 207–233.

[2] Sous-variétés d'une variété abélienne et points de torsion. In *Arithmetic and Geometry* (volume dedicated to Shafarevich), M. Artin, J. Tate, eds., Birkhäuser, 1983, 327–352.

[3] Around the Mordell conjecture for function fields and a conjecture of Serge Lang. In *Algebraic Geometry,* Lect. Notes in Math. 1016, Springer-Verlag, 1983, 1–19.

Rosen, M.

[1] Abelian varieties over $\mathbb{C}$. In *Arithmetic Geometry,* Cornell, Silverman, eds., Springer-Verlag, 1986, 79–102.

Rosenlicht, M.

[1] Some basic theorems on algebraic groups. *Amer. J. of Math.* **78** (1956), 401–443.

Roth, K.F.

[1] Rational approximations to algebraic numbers. *Mathematika* **2** (1955), 1–20.

Rubin, K.
[1] Tate–Shafarevich groups and L-functions of elliptic curves with complex multiplication. *Invent. Math.* **89** (1987), 527–560.
Schaefer, E.
[1] 2-Descent on the Jacobians of Hyperelliptic Curves. *J. of Number Theory* **51** (1995), 219–232.
Schanuel, S.
[1] Heights in number fields. *Bull. Soc. Math. France* **107** (1979), 433–449.
Schlickewei, H.
[1] The $\mathfrak{p}$-adic Thue–Siegel–Roth–Schmidt theorem. *Archiv. des Math.* **29** (1977), 267–270.
Schmidt, W.
[1] *Diophantine Approximation,* Lecture Notes in Math. 785, Springer-Verlag, Berlin, 1980.
[2] The subspace theorem in diophantine approximation. *Compositio Math.* **69** (1989), 121–173.
[3] Heights of points on subvarieties of $\mathbb{G}_m^s$. In *Number Theory 93–94,* S. David, eds., Cambridge University Press,
[4] Simultaneous approximation to algebraic number by rationals. *Acta Math.* **125** (1970), 189–201.
Serre, J.-P.
[1] *Groupes Algébriques et Corps de Classes,* Hermann, 1959.
[2] *Un cours d'arithmétique,* P.U.F., 1970; *A Course in Arithmetic,* Springer-Verlag, 1973.
[3] *Lectures on the Mordell–Weil Theorem,* Friedr. Vieweg & Sohn, 1989.
[4] *Corps Locaux,* Hermann, 1968; *Local Fields,* transl. by M. Greenberg, Springer-Verlag, 1979.
Shafarevich I.
[1] *Basic Algebraic Geometry,* Springer-Verlag, 1974.
[2] *Algebraic Surfaces,* Proc. Steklov Inst. Math. 75, 1965.
Shimura, G.
[1] *Introduction to the Arithmetic Theory of Automorphic Forms,* Princeton Univ. Press, 1971.
Shimura, G., Taniyama, Y.
[1] Complex Multiplication on Abelian Varieties. *Math. Soc. Japan* 1961.
Siegel, C.L.
[1] The integer solutions of the equation $y^2 = ax^n + bx^{n-1} + \cdots + k$. *J. London Math. Soc.* **1** (1926), 66–68.
[2] Über einege Anwendungen Diophantischer Approximationen. *Abh. Preuss. Akad. Wiss. Phys. Math. Kl.* (1929), 41–69.

Silverberg, A.

[1] Points of finite order on abelian varieties. In *p-adic Methods in Number Theory and Algebraic Geometry,* Adolphson, Sperber, Tretkoff, eds., Contemporary Math. 133, AMS, 1992, 175–193.

Silverman, J.H.

[1] *The Arithmetic of Elliptic Curves,* Springer-Verlag, 1986.

[2] *Advanced Topics in the Arithmetic of Elliptic Curves,* Springer-Verlag, 1994.

[3] A quantitative version of Siegel's theorem: integral points on elliptic curves and Catalan curves. *J. reine angew. Math.* **378** (1987), 60–100.

[4] Integral points on curves and surfaces. *Proc.* 15th *Journées Arith.*, 1987, Lect. Notes in Math. 1380, Springer-Verlag, 1989, 202–241.

[5] Lower bounds for height functions. *Duke Math. J.* **51** (1984), 395–403.

[6] Rational points on K3 surfaces: A new canonical height. *Invent. Math.* **105** (1991), 347–373.

[7] Computing heights on elliptic curves. *Math. Comp.* **51** (1988), 339–358.

[8] The theory of height functions. In *Arithmetic Geometry,* Cornell, Silverman, eds., Springer-Verlag, 1986, 151–166.

[9] Heights and elliptic curves. In *Arithmetic Geometry,* Cornell, Silverman, eds., Springer-Verlag, 1986, 253–265.

[10] Arithmetic distance functions and height functions in Diophantine geometry. *Math. Ann.* **279** (1987), 193–216.

[11] Heights and the specialization map for families of abelian varieties. *J. Reine angew. Math.* **342** (1983), 197–211.

Swinnerton-Dyer H.P.F.

[1] *Analytic Theory of Abelian Varieties,* London Math. Soc. L.N.S 14, Cambridge U. Press, 1974.

Szpiro, L.

[1] Séminaire sur les pinceaux de courbes de genre au moins 2. *Astérisque* **86** (1981).

[2] Séminaire sur les pinceaux arithmétiques : la conjecture de Mordell. *Astérisque* **127** (1985).

[3] Séminaire sur les pinceaux de courbes elliptiques. *Astérisque* **183** (1990).

[4] Sur les propriétés numériques du dualisant relatif d'une surface arithmétique. In *Grothendieck Festschrift III,* Birkhäuser, 1990, 229–246.

Szpiro, L., Ullmo, E., Zhang, S.

[1] Equidistribution des petits points. *Invent. Math.* **127** (1997), 337–347.

Tate, J.

[1] Residues and differentials on curves. *An. Sci. de l'E.N.S.* **1** (1968), 149–159.

[2] Arithmetic of elliptic curves. *Invent. Math.* **23** (1974), 179–206.

[3] On the conjecture of Birch and Swinnerton-Dyer. *Séminaire Bourbaki* exposé **306** (26 pages), February 1966.

[4] A review of non-archimedean elliptic functions. In *Elliptic Curves, Modular Forms, & Fermat's Last Theorem,* J. Coates and S.T. Yau, eds., International Press, Boston, 1995, 162–184.

Tenenbaum, G.

[1] Introduction à la théorie analytique et probabiliste des nombres. *Publ. Université de Nancy* **13** (1990).

Thunder, J.

[1] Asymptotic estimate for heights of algebraic subspaces. *Trans. Amer. Math. Soc.* **331** (1992), 395–424.

[2] Asymptotic estimate for rational points of bounded height on flag varieties. *Compo. Math.* **88** (1993), 155–183.

Thue, A.

[1] Über Annäherungswerte Algebraischer Zahlen. *J. Reine Angew. Math.* **135** (1909), 284–305.

Ueno, K.

[1] *Classification theory of algebraic varieties and complex spaces,* Lecture Notes Math 439, Springer-Verlag, 1975.

Ullmo, E.

[1] Positivité et discrétion des points algébriques sur les courbes. *Annals of Math.* **147** (1998), 167–179.

Vaughan, R.C., Wooley, T.

[1] On certain cubic forms and related equations. *Duke Math. J.* **80** (1995), 669–735.

Vojta, P.

[1] Siegel's theorem in the compact case. *Annals of Math.* **133** (1991), 509–548.

[2] Integral points on subvarieties of semi-abelian varieties. *Invent. Math.* **126** (1996), 133–181.

[3] *Diophantine approximation and value distribution theory,* Lect. Notes Math 1239, Springer-Verlag, 1987.

[4] Application of arithmetic algebraic geometry to diophantine approximations. Lect. Notes Math 1553, Springer-Verlag, 1993, 164–208.

van der Waerden, B.L.

[1] *Algebra* volume 1, Frederick Ungar, 1970.

Walker, R.J.

[1] *Algebraic Curves,* Princeton, 1950 (reissued Dover, 1962).

Weil, A.

[1] L'arithmétique sur les courbes algébriques. *Acta Math.* **52** (1928), 281–315.

[2] *Sur les Courbes Algébriques et les Variétés qui s'en Déduisent,* Hermann, Paris, 1948.

[3] *Variétés Abéliennes et Courbes Algébriques,* Hermann, 1948.

[4] Arithmetic on algebraic varieties. *Annals of Math.* **53** (1951), 412–444.

Wiles, A.

[1] Elliptic curves, modular forms and Fermat's last theorem. *Annals of Math.* **141** (1995), 443–551.

Zhang, S.

[1] Small points and adelic metrics. *J. Alg. Geom.* **4** (1995), 281–300.

[2] Equidistribution of small points on abelian varieties. *Annals of Math.* **147** (1998), 159–165.

[3] Positive line bundles on arithmetic surfaces. *Annals of Math.* **136** (1992), 569–587.

List of Notation

$\bar{k}$	an algebraic closure of k, 8
G_k	the Galois group of $\bar{k}$ over k, 8
$\mathrm{Gal}(\bar{k}/k)$	the Galois group of $\bar{k}$ over k, 8
$\mathbb{A}^n$	affine n-space, 9
$\mathbb{A}^n_k$	affine n-space over k, 9
$\mathbb{A}^n(k)$	the set of k-rational points of $\mathbb{A}^n$, 9
$Z(I)$	the set of zeros of the ideal I, 9
I_S	the ideal of polynomials vanishing on S, 9
$V(k)$	set of k rational points of V, 10
$I_{V,k}$	set of k-polynomials vanishing on V, 10
$\sqrt{I}$	the radical of the ideal I, 10
$\bar{k}[V]$	affine coordinate ring of V, 11
$\mathbb{P}^n$	projective n-space, 12
$\mathbb{P}^n(k)$	set of k-rational points of $\mathbb{P}^n$, 12
$k(P)$	field of definition of the point P, 12
$S(V)$	homogeneous coordinate ring of projective variety, 13
U_i	standard affine open subset of $\mathbb{P}^n$, 14
$\mathcal{O}(X)$	ring of regular functions on the variety X, 15
$\mathcal{O}_{x,X}$	local ring of the point x on the variety X, 15
$\mathcal{O}_x$	local ring of the point x on the variety X, 15
$\mathcal{O}_{Y,X}$	local ring of the variety X along the subvariety Y, 15
$\mathcal{M}_{Y,X}$	the maximal ideal of the local ring $\mathcal{O}_{Y,X}$, 15
$\bar{k}(X)$	function field of the variety X, 16
$\bar{k}(x_1,\ldots,x_n)$	the field of rational functions in n indeterminates, 16
$A_{\mathfrak{p}}$	localization at the prime ideal $\mathfrak{p}$, 17
$A_{(\mathfrak{p})}$	homogeneous localization at the prime ideal $\mathfrak{p}$, 17
$\mathrm{Frac}(A)$	ring of fractions of the ring A, 17
$\mathcal{M}_P$	ideal of functions vanishing at P, 18
$\deg(\phi)$	degree of the finite morphism ϕ, 19
Φ_d	d-uple embedding $\mathbb{P}^n \to \mathbb{P}^N$, 19
$S_{n,m}$	Segre map $\mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^N$, 19
π	linear projection $\mathbb{P}^n \to \mathbb{P}^r$, 19
$T_P(V)$	tangent space to the variety V at the point P, 24
f^*	induced map on cotangent space, 25
$\Omega^1[X]$	space of regular 1-forms on X, 26
$\bigwedge^r V$	space of r-linear skew-symmetric forms on V, 27

$\mathfrak{F}_{A,k}$	conductor of the abelian variety A/k, 453
$\sigma(E/k)$	the Szpiro ratio of the elliptic curve E/k, 454
Φ_E	modular parametrization $X_0(N) \to E$, 454
$f_E(z)$	weight 2 cusp form attached to elliptic curve E, 454
$\mathrm{Reg}(A/k)$	the canonical regulator of A/k, 459
$T_\ell(A)$	the Tate module of A, 460
$V_\ell(A)$	the Tate module of A tensored with $\mathbb{Q}$, 460
ρ_ℓ	the ℓ-adic representation of an abelian variety, 460
$L(A/\mathbb{Q}, s)$	the L-series of the abelian variety A, 461
$h_{\mathrm{Ar}}(P)$	the Arakelov height of P, 466
$\mathcal{A}(g,k,S)$	abelian varieties with good reduction outside S, 467
$\rho_{A,\ell}$	the ℓ-adic representation attached to A, 468
Φ_{mK_X}	rational map for pluricanonical divisor mK_X, 475
$g_m(X)$	the mth plurigenus of variety X, 475
$\kappa(X)$	the Kodaira dimension of the variety X, 475
Sp_X	the special subset of the variety X, 479
$m_S(D,P)$	arithmetic analogue of characteristic function, 483
$N_S(D,P)$	arithmetic analogue of counting function, 483
$\delta(\alpha)$	the arithmetic defect of α, 483
$Z_{\mathrm{geom}}(\varepsilon, E, D)$	geometric exceptional subset in Vojta conjecture, 484
$d_k(P)$	absolute logarithmic discriminant of P, 486
$N(X/k, D, B)$	counting function for k rational points on X, 487
$\mathrm{NS}_{\mathrm{eff}}(X)$	cone of effective divisor classes, 488
$\mathrm{NS}_+(X)$	cone of ample divisor classes, 488
$\alpha(D)$	Nevanlinna invariant of the divisor D, 488
$\partial\,\mathrm{NS}_{\mathrm{eff}}(X)$	the boundary of the effective cone, 488
$Z(U/k, D; s)$	height zeta function, 489

Index

Graduate Texts in Mathematics

(continued from page ii)

120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory. 2nd ed.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.

180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.
183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBAS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.
189 LAM. Lectures on Modules and Rings.
190 ESMONDE/MURTY. Problems in Algebraic Number Theory.
191 LANG. Fundamentals of Differential Geometry.
192 HIRSCH/LACOMBE. Elements of Functional Analysis.
193 COHEN. Advanced Topics in Computational Number Theory.
194 ENGEL/NAGEL. One-Parameter Semigroups for Linear Evolution Equations.
195 NATHANSON. Elementary Methods in Number Theory.
196 OSBORNE. Basic Homological Algebra.
197 EISENBUD/ HARRIS. The Geometry of Schemes.
198 ROBERT. A Course in p-adic Analysis.
199 HEDENMALM/KORENBLUM/ZHU. Theory of Bergman Spaces.
200 BAO/CHERN/SHEN. An Introduction to Riemann–Finsler Geometry.
201 HINDRY/SILVERMAN. Diophantine Geometry: An Introduction.